
TESTES EXATOS EM MODELOS HETEROSCEDÁSTICOS

WALDEMAR ARAÚJO DE SANTA CRUZ OLIVEIRA JÚNIOR

Orientadores: Prof. Dr. Francisco Cribari Neto e Profa. Dra. Silvia L.P. Ferrari

Área de concentração: Estatística Matemática

Dissertação submetida como requerimento parcial para obtenção do grau de Mestre em Estatística pela Universidade Federal de Pernambuco

Recife, maio de 2002

À memória do meu pai, Waldemar Araújo de Santa Cruz
Oliveira, que em toda sua vida se empenhou na formação
dos seus seis filhos.

Econometricians have found their Philosophers' Stone: it is called regression analysis and is used for transforming data into significant results. —David Hendry

Agradecimentos

A realização deste trabalho só foi possível com a graça de Deus e ajuda e cooperação de amigos no qual agradeço:

A meu pai, Waldemar Araújo de Santa Cruz Oliveira, por tamanho incentivo à minha formação, pelas inúmeras madrugadas em que me buscou na universidade e pelas alegres tardes de domingo que desde pequeno me propocionou no estádio da Ilha do Retiro.

À minha mãe Izabel Araújo de Santa Cruz Oliveira, principal responsável pela minha formação.

À minha esposa Cláudia Francisca de Amorim, pelo amor, compreensão e apoio.

A Larissa e Bianca, por serem a razão da minha vida.

A Belinha, Lua, João e Gabriel, por serem sobrinhos maravilhosos.

Às minhas cinco irmãs.

Aos Professores Francisco Cribari Neto e Silvia L. Ferrari, pela seriedade e apoio na orientação deste trabalho.

Aos Professores Manoel Lemos, Cristina Falcão Raposo, Cláudia Regina Oliveira de Paiva Lima, Klaus Leite Pinto Vasconcellos, André Toom, Manoel Raimundo Sena Junior, Sylvio José Pereira dos Santos e Isaac Xavier Junior, pela ajuda que sempre me deram.

A Valéria Bittencourt pela eficiência e simpatia.

Aos colegas Soraia, Franklin, Pedro, Nilton, Marcelo e Antônio, pela amizade e boa convivência no decorrer do curso.

A todos os meus amigos do Departamento de Estatística e do Departamento de Matemática desta universidade.

A Oscar, pela ajuda computacional.

À CAPES, pelo auxílio financeiro.

Resumo

Heteroscedasticidade é uma característica comumente encontrada em dados de corte transversal. Vários autores têm estudado o comportamento de estimadores consistentes da matriz de covariâncias do estimador de mínimos quadrados ordinários dos parâmetros lineares de regressão quando há heteroscedasticidade de forma desconhecida. Entre os estimadores propostos e estudados encontram-se aqueles conhecidos como HC0 (proposto por Halbert White em 1980), HC1, HC2 e HC3. Resultados de simulação em alguns artigos favorecem o estimador HC3 ou aproximações deste estimador; ver, por exemplo, MacKinnon & White (1985). Cribari-Neto & Galvão (2002), a partir dos resultados em Galvão (2000), generalizaram os resultados obtidos por Cribari-Neto, Ferrari & Cordeiro (2000), obtendo uma seqüência de estimadores ajustados por viés que pode ser inicializada em qualquer dos quatros estimadores listados acima. A presente dissertação utiliza integração numérica para obter resultados exatos sobre a qualidade da aproximação de primeira ordem usada em testes quase- t cujas estatísticas utilizam estimativas consistentes da variância do estimador de mínimos quadrados ordinários. Os resultados obtidos mostram que o teste que mais se beneficia de usar estimadores corrigidos por viés é aquele cuja estatística de teste é construída usando o estimador HC0. Adicionalmente, a utilização de estimativas da variância do tipo HC3 corrigidas por viés conduz a testes menos precisos, ao invés de conduzir a testes com menor distorção de tamanho. Por fim, mostra-se que a estratégia de inferência a ser preferida é a utilização de estimadores HC3 sem correção de viés.

Abstract

Cross sectional data with linear regression structure oftentimes display heteroskedastic behavior. A common practice is to estimate the linear parameters using ordinary least squares, and to obtain a consistent estimate for the covariance matrix of the ordinary least squares estimator. Several consistent estimators have been proposed and are widely known. The most commonly used ones are the HC0 (proposed by Halbert White in 1980), HC1, HC2 and HC3 estimators. Cribari–Neto, Ferrari & Cordeiro (2000) obtained a sequence of bias-adjusted HC0 estimators. Their results were generalized by Cribari–Neto & Galvão (2002), who derived bias corrections to a more general class of covariance matrix estimators, including the HC1, HC2 and HC3 estimators. The main goal of this thesis is to employ numerical integration methods to evaluate to what extent the use of bias-adjusted variance estimators in quasi- t test statistics leads to improved finite-sample inference. The results show that improvements in the quality of the first order approximation to the test are achieved when the bias-adjusted estimator is the White estimator (HC0). However, when the quasi- t test statistic is constructed using HC3 type estimators, bias adjustment do not result in associated tests with smaller size distortions, on the contrary. The results taken together suggest that the HC3 estimator should be preferred when performing quasi- t inference in linear regression models with heteroskedasticity of unknown form.

Índice

1. Introdução	1
1.1. Introdução	1
1.2. Regressão Linear	1
1.3. Plataforma Computacional.....	10
2. Estimadores Não-corrigidos e Corrigidos de Matrizes de Covariâncias	11
2.1. Introdução	11
2.2. Estimadores Consistentes	11
2.3. Seqüências de Estimadores Corrigidos	13
2.4. Estatísticas Quase- t	17
2.5. O Algoritmo de Imhof	21
3. Análise Exata de Testes por Integração Numérica	24
3.1. Introdução	24
3.2. Modelo e Teste	24
3.3. Resultados Numéricos	25
3.4. Conclusões	29
o Referências	58

Capítulo 1

Introdução

1.1. Introdução

Comumente nos deparamos com situações onde desejamos explicar determinados fenômenos através de outros. Por exemplo, poderíamos investigar a influência sobre a quantidade de votos recebida por um candidato a prefeito do montante de recursos gasto na campanha eleitoral e do tempo de propaganda eleitoral na televisão, ou ainda, como o índice de desemprego é afetado pelo índice da inflação e pelo total de investimento do governo realizado em educação. Nestes exemplos, estaríamos realizando uma análise de regressão.

Regressão linear é uma associação linear entre a média de uma variável de interesse, denominada variável dependente ou resposta, e um preditor que envolve parâmetros desconhecidos e variáveis auxiliares, a linearidade estando associada à relação entre a média da variável resposta e os parâmetros da estrutura de regressão. As variáveis auxiliares são chamadas de variáveis independentes ou explicativas ou mesmo de covariadas ou covariáveis. Inúmeras são as aplicações de um modelo linear em nosso cotidiano, por isso mesmo há uma enorme utilização do modelo linear de regressão em várias áreas, além da estatística, como sociologia, economia, biologia, etc.

Na modelagem de regressão, a variável resposta está sujeita a um erro aleatório. É comum supor que este erro tem média zero e variância constante. Contudo, a suposição de constância da variância do erro é muitas vezes violada em aplicações práticas. Quando isto ocorre dizemos que há heteroscedasticidade no modelo. Nesta perspectiva, iremos dar ênfase ao modelo heteroscedástico. Apresentaremos os estimadores de mínimos quadrados ordinário, generalizado e generalizado viável e o estimador de White da matriz de covariâncias do estimador de mínimos quadrados ordinários dos parâmetros de regressão, bem como alternativas a este estimador de covariâncias.

1.2. Regressão Linear

Considere uma amostra de tamanho n de pontos $(y_1, x_{11}, x_{12}, \dots, x_{1,p-1})$, $(y_2, x_{21}, x_{22}, \dots, x_{2,p-1})$, \dots , $(y_n, x_{n1}, x_{n2}, \dots, x_{n,p-1})$ e suponha que a variável y está relacionada com as variáveis x_i , $i = 1, \dots, p - 1$, através do modelo

$$y_t = \beta_0 + \beta_1 x_{t1} + \dots + \beta_{p-1} x_{tp-1} + u_t, \quad t = 1, \dots, n,$$

onde u_t é uma variável aleatória com média zero e variância σ_t^2 , chamada também de erro aleatório, e x_{ti} , para $t = 1, \dots, n$ e $i = 1, \dots, p-1$, é uma constante conhecida. Uma notação alternativa é

$$y = X\beta + u, \quad (1.1)$$

onde $y = (y_1, \dots, y_n)'$, $\beta = (\beta_0, \dots, \beta_{p-1})'$, $u = (u_1, \dots, u_n)'$ e X é uma matriz fixa de dimensão $n \times p$ ($p < n$), com posto igual a p , sendo que a primeira coluna desta matriz é um vetor de uns e o elemento $(t, i+1)$ é x_{ti} , com $i = 1, \dots, p-1$ e $t = 1, \dots, n$. Desta forma, podemos escrever o modelo linear de regressão como

$$\begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdot & \cdot & \cdot & x_{1,p-1} \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ 1 & x_{n1} & \cdot & \cdot & \cdot & x_{n,p-1} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \cdot \\ \beta_{p-1} \end{bmatrix} + \begin{bmatrix} u_1 \\ \cdot \\ \cdot \\ \cdot \\ u_n \end{bmatrix}.$$

Considere as seguintes suposições sobre o modelo em questão:

[S1] O modelo é de fato o correto, isto é, y é dado em (1.1).

[S2] $E[u_t] = 0$, $t = 1, \dots, n$, isto é, os erros têm média zero.

[S3] $\text{var}(u_t) = \sigma_t^2$ e $0 < \sigma_t^2 < \infty$, $t = 1, \dots, n$.

Sob homoscedasticidade, [S3] passa a ser

[S3'] $\text{var}(u_t) = \sigma^2$, $t = 1, \dots, n$ onde σ^2 é uma constante estritamente positiva.

[S4] $\text{cov}(u_t, u_s) = 0 \forall t \neq s$, isto é, os erros são não-correlacionados.

[S5] $\lim_{n \rightarrow \infty} n^{-1}(X'X) = Q$, onde Q é uma matriz finita positiva definida.

[S5'] $\lim_{n \rightarrow \infty} n^{-1}(X'\Omega X) = R$, onde R é uma matriz finita positiva definida.

[S6] $u \sim \mathcal{N}(0, \Omega)$, onde 0 é um vetor $n \times 1$ de zeros e $\Omega = \text{diag}\{\sigma_1^2, \dots, \sigma_n^2\}$.

Nosso objetivo é estimar o vetor de parâmetros β com base em n observações sobre as variáveis do modelo. Seja \hat{y} o vetor de valores ajustados, isto é, $\hat{y} = X\hat{\beta}$, onde $\hat{\beta}$ é o estimador de mínimos quadrados ordinários de β dado em (1.2) abaixo. O vetor de resíduos de mínimos quadrados é então definido como

$$\hat{u} = y - \hat{y} = y - X\hat{\beta}.$$

A maneira mais comumente utilizada para se estimar β é minimizando a soma dos quadrados dos erros, ou seja, encontrando o valor de β que minimiza a função

$$S^2(\beta) = \sum_{t=1}^n u_t^2 = u'u = (y - X\beta)'(y - X\beta).$$

É fácil mostrar que o mínimo de $S^2(\beta)$ ocorre em

$$\hat{\beta} = (X'X)^{-1}X'y. \quad (1.2)$$

Este é o estimador de mínimos quadrados ordinários (EMQO) de β .

Este estimador possui propriedades desejáveis, pois, admitindo-se que estamos estimando o modelo correto, isto é, considerando [S1]¹ e que a matriz X é fixa, obtemos as seguintes propriedades:

i) Sob [S2], $\hat{\beta}$ é um estimador não-viesado de β .

Dizemos que um estimador é não viesado se seu valor esperado é igual ao parâmetro sendo estimado, para todos os valores do parâmetro no espaço paramétrico. Ou seja, se $E[\hat{\beta}] = \beta$, para todo $\beta \in \mathbb{R}^p$, então $\hat{\beta}$ é um estimador não-viesado para β .

Para estabelecer esta propriedade, observemos que

$$\begin{aligned} \hat{\beta} &= (X'X)^{-1}X'y \\ &= (X'X)^{-1}X'(X\beta + u) \\ &= \beta + (X'X)^{-1}X'u. \end{aligned}$$

Então, $E[\hat{\beta}] = \beta + (X'X)^{-1}X'E[u]$, este último termo sendo igual a zero uma vez que os erros têm média zero. Usamos o fato de X ser uma matriz fixa e portanto $E[(X'X)^{-1}X'u] = (X'X)^{-1}X'E[u]$, para concluir que $E[\hat{\beta}] = \beta$.

ii) Se considerarmos [S2] e [S5] temos que $\hat{\beta}$ é um estimador consistente para β .

Uma seqüência de estimadores $\hat{\theta}_n$ de um parâmetro θ é consistente se para todo $\epsilon > 0$ $\lim_{n \rightarrow \infty} P(\|\hat{\theta}_n - \theta\| > \epsilon) = 0$, ou seja, se $\hat{\theta}_n$ converge em probabilidade para θ . Dizemos neste caso que há convergência fraca (White, 1984, p. 22). Convergência forte se dá quando a convergência é quase certa.

As notações mais utilizadas para denotar que $\hat{\beta}_n$ converge em probabilidade para β quando n tende para infinito são $\hat{\beta}_n \xrightarrow{p} \beta$ ou $\text{plim}(\hat{\beta}_n) = \beta$.

Antes de verificarmos a consistência de $\hat{\beta}_n$, notemos que $\text{plim}(\|\frac{X'u}{n}\|) = 0$, pois pela desigualdade de Chebyshev

$$P\left(\left\|\frac{X'u}{n}\right\| \geq \epsilon\right) \leq \frac{E[\|X'u\|]}{\epsilon n} \rightarrow 0 \quad \text{quando } n \rightarrow \infty.$$

Logo,

$$\text{plim}\left(\left\|\frac{X'u}{n}\right\|\right) = 0.$$

¹ Sem a Suposição [S1] muitos de nossos resultados não seriam verdadeiros.

Para este resultado nós utilizamos [S5']:

$$\begin{aligned}
P\left(\left\|\frac{X'u}{n}\right\| \geq \epsilon\right) &= P\left(\left\|\frac{X'u}{n}\right\|^2 \geq \epsilon^2\right) = P(\|X'u\|^2 \geq n^2\epsilon^2) \leq \frac{1}{n^2\epsilon^2} E[\|X'u\|^2] \\
&= \frac{1}{n^2\epsilon^2} E[u'X X'u] = \frac{1}{n^2\epsilon^2} E[\text{tr}(u'X X'u)] = \frac{1}{n^2\epsilon^2} \text{tr}\{E[XX'u u']\} \\
&= \frac{1}{n^2\epsilon^2} \text{tr}\{XX'\Omega\} = \frac{\text{tr}\{X'\Omega X\}}{n^2\epsilon^2} = \frac{\text{tr}\{X'\Omega X/n\}}{n\epsilon^2}.
\end{aligned}$$

Do teorema de Slutsky, que afirma que o limite de uma soma (ou produto) de uma seqüência estocástica é a soma (ou produto) dos limites desta seqüência, temos que

$$\begin{aligned}
\text{plim}(\widehat{\beta}) &= \text{plim}[\beta + (X'X)^{-1}X'u] \\
&= \text{plim}(\beta) + \text{plim}[n(X'X)^{-1}] \text{plim}\left(\frac{X'u}{n}\right) \\
&= \beta + Q^{-1}0 = \beta,
\end{aligned}$$

o que demonstra a consistência de $\widehat{\beta}$.

iii) *Dadas as Suposições S[2], S[3'] e S[4], o EMQO é o estimador de menor variância dentre todos os estimadores lineares² e não-viesados. Ou seja, o EMQO é dito ser o melhor estimador linear e não-viesado de β .*

Um estimador $\widehat{\beta}$ não-viesado para β é dito ser melhor que um outro estimador $\widetilde{\beta}$, também não-viesado para β , se a diferença entre as matrizes de covariâncias $\widetilde{\psi}_{\widetilde{\beta}}$ de $\widetilde{\beta}$ e $\psi_{\widehat{\beta}}$ de $\widehat{\beta}$ é positiva semi-definida.³

Para calcular a matriz de covariâncias de $\widehat{\beta}$ usamos o fato de que $\widehat{\beta} - \beta = (X'X)^{-1}X'u$, obtendo assim

$$\begin{aligned}
\psi_{\widehat{\beta}} &= \text{cov}(\widehat{\beta}) \\
&= E[(\widehat{\beta} - E[\widehat{\beta}])(\widehat{\beta} - E[\widehat{\beta}])'] \\
&= E[(\widehat{\beta} - \beta)(\widehat{\beta} - \beta)'] \\
&= E[(X'X)^{-1}X'u u'X(X'X)^{-1}] \\
&= (X'X)^{-1}X'E[uu']X(X'X)^{-1}.
\end{aligned}$$

Desta forma, chegamos à expressão para a matriz de covariâncias de $\widehat{\beta}$:

$$\psi_{\widehat{\beta}} = (X'X)^{-1}X'\Omega X(X'X)^{-1}, \tag{1.3}$$

² Se $\widehat{\beta} = Ay$, onde A é uma matriz de ordem $p \times n$ e y o vetor de respostas, então $\widehat{\beta}$ é linear.

³ Uma matriz quadrada A de ordem n é dita ser positiva semi-definida se para todo vetor n -dimensional x temos que $x'Ax \geq 0$. Neste caso escrevemos $A \geq 0$. No caso em que a desigualdade é sempre estrita, $x'Ax > 0, \forall x \neq 0$, dizemos que A é positiva definida e escrevemos $A > 0$.

onde $\Omega = E[uu'] = \text{cov}(u)$.

Um modelo de regressão é dito ser homoscedástico se as variâncias dos erros são constantes. Então, sob homoscedasticidade ([S3']) e não-autocorrelação ([S4]) a matriz $\psi_{\hat{\beta}}$ em (1.3) passa a ter a forma

$$\psi_{\hat{\beta}} = \sigma^2(X'X)^{-1},$$

pois aqui $\Omega = \sigma^2 I_n$, onde I_n é a matriz identidade $n \times n$.

Iremos concluir que o EMQO sob as Suposições [S1], [S2], [S3'] e [S4] é o melhor estimador linear e não-viesado de β . Isto se encontra descrito no próximo teorema.

Teorema 1. (Teorema de Gauss–Markov) *Sob as Suposições [S1], [S2], [S3'] e [S4], dentre todos os estimadores lineares e não-viesados de β , o EMQO possui variância mínima. Em outras palavras, se $\tilde{\beta} = Ay$ e $E[\tilde{\beta}] = \beta$, então $\text{var}(c'\tilde{\beta}) \geq \text{var}(c'\hat{\beta})$, $\forall c \in \mathbb{R}^p$, onde A é uma matriz fixa $p \times n$.*

Para provar este resultado, assuma que β^* é um estimador linear para β , isto é, tem a forma $\beta^* = [(X'X)^{-1}X' + A]y$, onde A é qualquer matriz fixa $p \times n$. (Se $A = 0$, então $\beta^* = \hat{\beta}$.) Dado [S1],

$$\begin{aligned}\beta^* &= [(X'X)^{-1}X' + A](X\beta + u) \\ &= \beta + (X'X)^{-1}X'u + AX\beta + Au.\end{aligned}$$

Com a Suposição [S2] temos que $E[\beta^*] = \beta + AX\beta$. Para que β^* seja não-viesado é necessário que $AX = 0$. Logo, $\beta^* - \beta = [A + (X'X)^{-1}X']u$. Sob as hipóteses [S3'], [S4] e dado que $AX = 0$, a matriz de covariâncias de β^* é dada por

$$\begin{aligned}\text{cov}(\beta^*) &= E[(\beta^* - \beta)(\beta^* - \beta)'] \\ &= (A + (X'X)^{-1}X')E[uu'](A' + X(X'X)^{-1}) \\ &= (A + (X'X)^{-1}X')\sigma^2 I_n(A' + X(X'X)^{-1}) \\ &= \sigma^2(AA' + (X'X)^{-1}X'A' + AX(X'X)^{-1} + (X'X)^{-1}X'X(X'X)^{-1}) \\ &= \sigma^2(AA' + (X'X)^{-1}).\end{aligned}$$

Portanto,

$$\text{cov}(\beta^*) - \text{cov}(\hat{\beta}) = \sigma^2(AA' + (X'X)^{-1}) - \sigma^2(X'X)^{-1} = \sigma^2 AA' \geq 0.$$

Ou seja, esta matriz de diferença de covariâncias é positiva semi-definida.

O que este teorema nos afirma é que o EMQO é o melhor estimador linear e não-viesado (BLUE – ‘Best Linear Unbiased Estimator’) de β . Note que a suposição de homoscedasticidade [S3'] é necessária para este resultado e que ele não permanece válido quando há heteroscedasticidade, ou seja, quando as variâncias dos erros não são constantes. Para verificar este fato, comecemos definindo um “novo” estimador linear e não-viesado para β ,

no modelo heteroscedástico, isto é, supondo [S3] ao invés de [S3']. Seja $y = X\beta + u$, com $\text{cov}(u) = \Omega = \text{diag}\{\sigma_1^2, \dots, \sigma_n^2\} \neq \sigma^2 I_n$. Como Ω é uma matriz positiva definida, sua inversa existe e ela pode ser escrita como $\Omega^{-1} = \Omega^{-1/2}\Omega^{-1/2}$. Podemos multiplicar $\Omega^{-1/2}$ pela esquerda no modelo (1.1), isto é, $\Omega^{-1/2}y = \Omega^{-1/2}X\beta + \Omega^{-1/2}u$, obtendo

$$y^* = X^*\beta + u^*,$$

com $y^* = \Omega^{-1/2}y$, $X^* = \Omega^{-1/2}X$ e $u^* = \Omega^{-1/2}u$. Então, $E[u^*] = E[\Omega^{-1/2}u] = \Omega^{-1/2}E[u] = 0$ e $\text{cov}(u^*) = \Omega^{-1/2}\text{cov}(u)\Omega^{-1/2} = I_n$. Ou seja, este novo modelo, que é homoscedástico, satisfaz todas suposições do Teorema 1. Logo, o estimador MQO de β neste modelo transformado é o melhor estimador linear e não-viesado. Minimizando a soma do quadrados dos erros obtemos $\hat{\beta}_G = (X^{*\prime}X^*)^{-1}X^{*\prime}y^*$, ou seja,

$$\begin{aligned}\hat{\beta}_G &= (X'\Omega^{-1/2}\Omega^{-1/2}X)^{-1}X'\Omega^{-1/2}\Omega^{-1/2}y \\ &= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y.\end{aligned}$$

Como era de se esperar, este estimador é linear e não-viesado. De fato,

$$\begin{aligned}E[\hat{\beta}_G] &= E[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y] \\ &= E[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}X\beta + (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}u] \\ &= \beta.\end{aligned}$$

Sua matriz de covariâncias é dada por

$$\begin{aligned}\psi_{\hat{\beta}_G} &= \text{cov}(\hat{\beta}_G) \\ &= (X^{*\prime}X^*)^{-1}X^{*\prime}\Omega^*X^*(X^{*\prime}X^*)^{-1} \\ &= (X'\Omega^{-1/2}\Omega^{-1/2}X)^{-1}X'\Omega^{-1/2}I_n\Omega^{-1/2}X(X'\Omega^{-1/2}\Omega^{-1/2}X)^{-1}, \\ &= (X'\Omega^{-1}X)^{-1}.\end{aligned}$$

O que fizemos foi usar o EMQO para estimar β no modelo transformado $y^* = X^*\beta + u^*$. Chamaremos este estimador de estimador de mínimos quadrados generalizados (EMQG). Note que se o modelo for homoscedástico o EMQG e o EMQO coincidem, pois $\Omega = \sigma^2 I_n$ e, portanto,

$$\begin{aligned}\hat{\beta}_G &= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}u \\ &= \sigma^2(X'X)^{-1}X'(\sigma^2)^{-1}u \\ &= (X'X)^{-1}X'u \\ &= \hat{\beta}.\end{aligned}$$

Evidentemente, suas matrizes de covariâncias também coincidem já que

$$\begin{aligned}\psi_{\widehat{\beta}_G} &= (X'\Omega^{-1}X)^{-1} \\ &= \sigma^2(X'X)^{-1} \\ &= \psi_{\widehat{\beta}}.\end{aligned}$$

O estimador $\widehat{\beta}$ no modelo heteroscedástico permanece linear e não-viesado, porém não terá variância mínima. De fato,

$$\begin{aligned}\psi_{\widehat{\beta}} - \psi_{\widehat{\beta}_G} &= (X'X)^{-1}X'\Omega X(X'X)^{-1} - (X'\Omega^{-1}X)^{-1} \\ &= [(X'X)^{-1}X' - (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}]\Omega[X(X'X)^{-1} - \Omega^{-1}X(X'\Omega^{-1}X)^{-1}] \\ &= A\Omega A' \geq 0,\end{aligned}$$

onde $A = (X'X)^{-1}X' - (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}$. A desigualdade $A\Omega A' \geq 0$ acontece porque Ω é diagonal com todos elementos positivos; então para qualquer vetor x de dimensão $1 \times n$, teremos que $x\Omega x' \geq 0$, mas xA também tem dimensão $1 \times n$, logo a desigualdade está confirmada.

Davidson e MacKinnon (1993, p. 549) ilustram a utilização dos dois estimadores em um modelo heteroscedástico através de um estudo de Monte Carlo. Estes autores consideraram o modelo $y_t = \beta_1 + \beta_2 x_t + u_t$, onde $\beta_1 = 1$, $\beta_2 = 1$ e $u_t \sim \mathcal{N}(0, x_t^\alpha)$, sendo α um valor dado e conhecido, e calcularam as variâncias de $\widehat{\beta}_1$ e $\widehat{\beta}_2$ de três formas, a saber: $\widehat{\sigma}^2(X'X)^{-1}$ ('MQO'), $(X'X)^{-1}X'\Omega X(X'X)^{-1}$ ('MQO_{het}') e $(X'\Omega^{-1}X)^{-1}$ ('MQG'), onde $\widehat{\sigma}^2$ é o estimador usual $\widehat{u}'\widehat{u}/(n-2)$ e $\Omega = \text{diag}\{x_1^\alpha, \dots, x_n^\alpha\}$; utilizando 20,000 (vinte mil) réplicas de Monte Carlo os autores estimaram numericamente estas variâncias. Estes resultados estão apresentados na Tabela 1.1 para diversos valores de α , onde são apresentados os valores médios dos erros padrão nas 20,000 réplicas de Monte Carlo.

Tabela 1.1. Ensaio de Monte Carlo realizado por Davidson e MacKinnon.

α	Erro padrão de $\widehat{\beta}_1$			Erro padrão de $\widehat{\beta}_2$		
	MQO	MQO _{het}	MQG	MQO	MQO _{het}	MQG
0.5	0.164	0.134	0.110	0.285	0.277	0.243
1.0	0.142	0.101	0.048	0.246	0.247	0.173
2.0	0.116	0.074	0.007	0.200	0.220	0.109
3.0	0.100	0.064	0.001	0.173	0.206	0.056
4.0	0.089	0.059	0.000	0.154	0.195	0.017

Observação: MQO_{het} significa MQO levando em conta a heteroscedasticidade.

Notamos uma superioridade do EMQG em relação ao EMQO, independente de levarmos em consideração a heteroscedasticidade.⁴ Em outras palavras, o estimador MQO apesar de linear e não-viesado não é mais eficiente, salvo quando o modelo é homoscedástico, pois, como foi visto, sem a presença de heteroscedasticidade as matrizes $\widehat{\psi}_{\widehat{\beta}}$ e $\widehat{\psi}_{\widehat{\beta}_G}$ coincidem.

Todavia, o problema central é que a matriz de covariâncias dos erros Ω em geral é desconhecida, tornando o EMQG inviável. Uma solução para este problema é substituir Ω por algum estimador consistente $\widehat{\Omega}$, conduzindo assim ao estimador de mínimos quadrados generalizados viável (EMQGV)

$$\widehat{\widehat{\beta}} = (X'\widehat{\Omega}^{-1}X)^{-1}X'\widehat{\Omega}^{-1}y.$$

Uma maneira de estimar Ω é fazendo suposições sobre a forma do comportamento dos seus elementos. Por exemplo, se encontrarmos suspeitas de heteroscedasticidade multiplicativa podemos utilizar

$$\sigma_t^2 = \exp(z_t'\alpha), \quad t = 1, \dots, n, \quad (1.4)$$

proposto por Harvey (1976), onde cada z_t , de dimensão $q \times 1$, é uma função da linha t da matriz X em \mathbb{R}^q e α é um vetor de parâmetros desconhecidos de dimensão $q \times 1$. Quando o modelo tem apenas uma variável dependente e esta é positiva, se definirmos $z_t = \ln(x_t)$ e $\alpha = \lambda$, onde x_t é a observação t e λ uma constante desconhecida, teremos neste caso particular $\sigma_t^2 = x_t^\lambda$, que é uma suposição comumente utilizada sobre a forma da variância. Após definir a forma de z_t , encontramos $\widehat{\Omega}$ substituindo σ_t^2 por \widehat{u}_t^2 em (1.4), obtendo assim os pontos

$$\widehat{u}_t^2 = \exp(z_t'\alpha) \times \nu_t,$$

ou ainda

$$\ln(\widehat{u}_t^2) = \alpha_1 z_{t1} + \dots + \alpha_q z_{tq} + \ln(\nu_t),$$

sendo ν um vetor de erros de média zero, e estimamos α pelo método de mínimos quadrados. O exemplo descrito a seguir ilustra tal procedimento.

Utilizando o pacote estatístico GRETL, disponível gratuitamente na página <http://gretl.sourceforge.net>, retiramos uma amostra de tamanho 36 do banco de dados de William Greene sobre demanda por gasolina, este sendo um dos bancos de dados que acompanha o programa (Greene, 1997). Aqui relacionamos o consumo total de gasolina nos Estados Unidos, definido como a variável y , com o preço da gasolina, o preço do transporte público e o total da população em milhões: x_1 , x_2 e x_3 , respectivamente. O modelo utilizado foi $y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \beta_3 x_{t3} + u_t$, $t = 1, \dots, 36$. As principais etapas para a obtenção de $\widehat{\beta}$ estão enumeradas a seguir.

1) Fazemos um teste de heteroscedasticidade, pois só iremos usar o EMQGV caso acreditamos que o modelo não é homoscedástico.

⁴ Quando não levamos em consideração a heteroscedasticidade o EMQO reduz-se à expressão $\widehat{\sigma}^2(X'X)^{-1}$.

Os erros padrão foram calculados a partir da raiz quadrada dos elementos diagonais da matriz $\widehat{\Psi}$, que no caso do EMQO é dada por

$$\widehat{\Psi} = \widehat{\sigma}^2(X'X)^{-1},$$

onde $\widehat{\sigma}^2$ é o estimador usual de mínimos quadrados. Para o EMQGV é dada por

$$\widehat{\Psi} = \widehat{\sigma}^2(X'\widehat{\Omega}^{-1}X)^{-1},$$

com $\widehat{\sigma}^2 = (y - X(X'\widehat{\Omega}^{-1}X)^{-1}X'\widehat{\Omega}^{-1}y)'\widehat{\Omega}^{-1}(y - X(X'\widehat{\Omega}^{-1}X)^{-1}X'\widehat{\Omega}^{-1}y)/33$; ver Judge et al. (1988, p. 369). A estatística t é utilizada para testar a hipótese $\mathcal{H}_0 : c'\beta = 0$ versus $\mathcal{H}_0 : c'\beta \neq 0$. O teste de heteroscedasticidade de White rejeita a hipótese nula de variâncias constantes ao nível de significância de 10%, apresentando valor p de 0.0581. Comparando as Tabelas 1.2 e 1.3 notamos diferenças nas estimativas dos parâmetros e em seus erros padrão.

Um aspecto indesejável do método de MQGV é que ele requer que o usuário faça suposições sobre a forma funcional do mecanismo heteroscedástico. A maneira mais usual de evitar suposições paramétricas sobre o comportamento funcional da variância é utilizar o EMQO, que permanece não-viesado e consistente quando há heteroscedasticidade no modelo, e utilizar um estimador $\widehat{\Omega}$ tal que $X'\widehat{\Omega}X$ seja consistente para $X'\Omega X$. Esta abordagem será utilizada na presente dissertação.

White (1980) propôs um estimador consistente para a matriz de covariâncias do EMQO e posteriormente versões modificadas desse estimador foram sugeridas. Cribari–Neto, Ferrari & Cordeiro (2000) propuseram uma seqüência de estimadores corrigidos a partir do estimador de White cujo viés converge mais rapidamente para zero. Eles construíram o primeiro elemento $\widehat{\psi}^{(1)}$ desta seqüência subtraindo o viés estimado do estimador de White deste estimador, o segundo elemento subtraindo o viés estimado de $\widehat{\psi}^{(1)}$ dele mesmo, e assim sucessivamente. Cribari–Neto & Galvão (2002), a partir dos resultados da dissertação de mestrado de Nila Galvão (Galvão, 2000), generalizaram este resultado para cobrir algumas modificações do estimador de White.

1.3. Plataforma Computacional

As avaliações numéricas realizadas ao longo da presente dissertação utilizaram a linguagem matricial de programação `0x`; ver Cribari–Neto (1997) e Doornik (2001). Trata-se de uma linguagem de programação distribuída gratuitamente para uso acadêmico e disponível em <http://www.nuff.ox.ac.uk/Users/Doornik>. Os gráficos apresentados foram confeccionados utilizando o ambiente de programação, análise de dados e gráficos `R`, que se encontra disponível gratuitamente em <http://www.r-project.org>. Por fim, a dissertação foi datilografada usando o sistema de tipografia (Plain) `TEX`, também distribuído de forma gratuita; ver Knuth (1986).

Capítulo 2

Estimadores Não-corrigidos e Corrigidos de Matrizes de Covariâncias

2.1. Introdução

Em um modelo linear heteroscedástico no qual a forma da heteroscedasticidade é conhecida podemos utilizar o estimador de mínimos quadrados generalizados para estimar o parâmetro β , como visto no capítulo anterior. Caso o mecanismo de heteroscedasticidade não seja conhecido, este estimador não é viável e uma alternativa seria usar o estimador MQGV. Porém, para tanto seria necessário fazer suposições sobre a forma funcional das variâncias dos erros, o que em geral se deseja evitar. Um enfoque alternativo é utilizar o estimador de mínimos quadrados ordinários para estimar o parâmetro β e um estimador consistente para sua matriz de covariâncias. Neste capítulo apresentaremos alguns estimadores consistentes para esta matriz. Iniciaremos com o estimador de White (HC0), demonstrando sua consistência. Apresentamos também três alternativas para este estimador, a saber: os estimadores HC1, sugerido por Hinkley (1977), HC2, proposto por Horn, Horn & Ducan (1975), e HC3, descrito em Davidson & MacKinnon (1993). Apresentaremos também a seqüência de estimadores corrigidos para o estimador de White desenvolvida por Cribari-Neto, Ferrari & Cordeiro (2000) e generalizada por Cribari-Neto & Galvão (2002); ver também Galvão (2000). Esta seqüência gera estimadores com vieses cada vez menores em ordem. O objetivo central da presente dissertação é avaliar como a redução de viés influencia o desempenho dos testes associados.

2.2. Estimadores Consistentes

Quando se suspeita da presença de heteroscedasticidade, uma das maneiras mais usadas para estimar o parâmetro β é utilizar o EMQO e algum estimador consistente para sua matriz de covariâncias $\psi_{\hat{\beta}}$. Halbert White, em 1980, propôs um estimador consistente. Ao invés de tentar estimar Ω , que tem n elementos, ele buscou estimar $X'\Omega X$, que tem $p(p+1)/2$ elementos distintos, independentemente do tamanho amostral n , da seguinte forma:

$$\text{HC0} = (X'X)^{-1}X'\hat{\Omega}X(X'X)^{-1},$$

onde $\hat{\Omega} = \text{diag}\{\hat{u}_1^2, \dots, \hat{u}_n^2\}$. Aqui, \hat{u}_t é o t -ésimo resíduo, i.e., é o t -ésimo elemento de $\hat{u} = y - X\hat{\beta}$. White (1980) mostrou que $X'\hat{\Omega}X \xrightarrow{q.c.} X'\Omega X$.

O estimador de White, também conhecido como HC0, apesar de consistente, tende a subestimar a variância do EMQO em amostras finitas. Uma correção para esta tendência,

sugerida por Hinkley (1977), consiste em utilizar o número de graus de liberdade como correção, similarmente ao que é feito com $\hat{\sigma}^2$, obtendo-se assim o estimador

$$\text{HC1} = \frac{n}{(n-p)} (X'X)^{-1} X' \hat{\Omega} X (X'X)^{-1}.$$

Horn, Horn & Ducan (1975) definiram o estimador HC2 também com o intuito de amenizar a subestimação dos erros pelos resíduos. Temos que, sob homoscedasticidade,

$$\begin{aligned} E[\widehat{u}\widehat{u}'] &= E[(y - \widehat{y})(y - \widehat{y})'] \\ &= E[(I - H)y((I - H)y)'] \\ &= E[(I - H)yy'(I - H)'] \\ &= ME[yy']M' \\ &= \sigma^2 M, \end{aligned}$$

onde $M = (I - H)$ e $H = X(X'X)^{-1}X'$. Note que acima usamos o fato de que a matriz M é simétrica e idempotente. Assim,

$$\begin{aligned} E[\widehat{u}_t^2] &= \sigma^2 m_t \\ &= \sigma^2(1 - h_t), \end{aligned}$$

onde m_t e h_t são os t -ésimos elementos diagonais de M e H , respectivamente. É intuitivo, portanto, definir o estimador HC2 da seguinte forma:

$$\text{HC2} = (X'X)^{-1} X' \widehat{\Omega}_2 X (X'X)^{-1},$$

onde

$$\widehat{\Omega}_2 = \text{diag} \left\{ \frac{\widehat{u}_1^2}{(1 - h_1)}, \dots, \frac{\widehat{u}_n^2}{(1 - h_n)} \right\}.$$

Sob homoscedasticidade, este estimador é não-viesado, pois

$$\begin{aligned} E[\text{HC2}] &= E[(X'X)^{-1} X' \widehat{\Omega}_2 X (X'X)^{-1}] \\ &= (X'X)^{-1} X' E[\widehat{\Omega}_2] X (X'X)^{-1} \\ &= (X'X)^{-1} X' \text{diag} \left\{ \frac{\sigma^2(1 - h_1)}{1 - h_1}, \dots, \frac{\sigma^2(1 - h_n)}{1 - h_n} \right\} X (X'X)^{-1} \\ &= \sigma^2 (X'X)^{-1} \\ &= \psi_{\widehat{\beta}}. \end{aligned}$$

Podemos ainda utilizar a técnica jackknife para obter uma alternativa ao estimador HC0. Como afirma Galvão (2000, p. 11), “a idéia básica desta técnica é recalcular n vezes as estimativas de mínimos quadrados ordinários para o vetor β , cada vez retirando uma

das observações, e então usar a variância das estimativas obtidas como uma estimativa da variância do estimador de MQO.” O estimador HC3, descrito em Davidson & MacKinnon (1993, p. 554) fornece uma aproximação para o estimador jackknife. Aqui, o t -ésimo elemento diagonal de $\widehat{\Omega}_3$ é dado por $\widehat{u}_t^2/(1 - h_t)^2$, ou seja,

$$\text{HC3} = (X'X)^{-1}X'\widehat{\Omega}_3X(X'X)^{-1},$$

onde

$$\widehat{\Omega}_3 = \text{diag} \left\{ \frac{\widehat{u}_1^2}{(1 - h_1)^2}, \dots, \frac{\widehat{u}_n^2}{(1 - h_n)^2} \right\}.$$

Em suma, apresentamos acima quatro estimadores alternativos para a matriz de covariâncias do estimador de MQO de β . São eles:

$$\widehat{\psi}_i = P\widehat{\Omega}_iP', \quad i = 0, \dots, 3,$$

onde $\widehat{\Omega}_i = D_i\widehat{\Omega}$, $\widehat{\Omega} = \text{diag}\{\widehat{u}_1^2, \dots, \widehat{u}_n^2\}$, $P = (X'X)^{-1}X'$ e D_i é uma matriz diagonal $n \times n$ definida, conforme o estimador considerado, como

$$\text{HC0} : D_0 = I_n, \tag{2.2}$$

$$\text{HC1} : D_1 = \frac{n}{n - p}I_n, \tag{2.3}$$

$$\text{HC2} : D_2 = \text{diag} \left\{ \frac{1}{1 - h_1}, \dots, \frac{1}{1 - h_n} \right\}, \tag{2.4}$$

$$\text{HC3} : D_3 = \text{diag} \left\{ \frac{1}{(1 - h_1)^2}, \dots, \frac{1}{(1 - h_n)^2} \right\}. \tag{2.5}$$

2.3. Seqüências de Estimadores Corrigidos

Cribari-Neto, Ferrari & Cordeiro (2000) definiram uma seqüência de estimadores corrigidos para a matriz de covariâncias de $\widehat{\beta}$ partindo do estimador HC0. O k -ésimo elemento desta seqüência é o $(k - 1)$ -ésimo elemento menos o seu viés estimado e o elemento inicial desta seqüência é o estimador HC0. A construção para o termo geral da seqüência foi facilitada com a definição de uma função recursiva $M^{(k)}(\Omega)$, estudada abaixo. Cribari-Neto & Galvão (2002) generalizaram esta seqüência iniciando-a a partir de qualquer um dos quatro estimadores em (2.2)–(2.5); ver também Galvão (2000). Na apresentação e estudo desta seqüência demonstramos, através de indução matemática, todas as expressões utilizadas.

A idéia central é definir uma seqüência de estimadores corrigidos. Ou seja, dado um estimador para $\psi_{\widehat{\beta}}$ definiremos um outro que é o próprio estimador menos o seu viés estimado. Este novo estimador, por sua vez, apresenta um novo viés. Repetimos este processo k vezes, obtendo assim $\widehat{\psi}^{(k)}$, que é o estimador corrigido k vezes para a matriz de covariâncias do estimador de MQO para o parâmetro β .

Considere a seguinte função recursiva de uma matriz A de dimensão $n \times n$:

$$M^{(k+1)} = M^{(1)}(M^{(k)}(A)), \quad k = 0, 1, 2, \dots,$$

onde $M^{(0)}(A) = A$, $M^{(1)}(A) = \text{diag}\{HA(H - 2I)\}$ e $H = X(X'X)^{-1}X'$. Esta função possui três propriedades importantes, que enunciamos e demonstramos a seguir via indução matemática.

[P1] $M^{(k)}(A) + M^{(k)}(B) = M^{(k)}(A + B)$.

Para $k = 0$, temos $M^{(0)}(A) + M^{(0)}(B) = A + B = M^{(0)}(A + B)$. Suponha que esta regra vale para $k > 0$ iterações. Então,

$$\begin{aligned} M^{(k+1)}(A) + M^{(k+1)}(B) &= M^{(1)}(M^{(k)}(A)) + M^{(1)}(M^{(k)}(B)) \\ &= \text{diag}\{HM^{(k)}(A)(H - 2I)\} + \text{diag}\{HM^{(k)}(B)(H - 2I)\} \\ &= \text{diag}\{H(M^{(k)}(A) + M^{(k)}(B))(H - 2I)\} \\ &= \text{diag}\{H(M^{(k)}(A + B))(H - 2I)\} \\ &= M^{(1)}(M^{(k)}(A + B)) \\ &= M^{(k+1)}(A + B). \end{aligned}$$

[P2] $M^{(k)}(M^{(1)}(A)) = M^{(k+1)}(A)$.

Para $k = 0$ o resultado é trivial. Suponha que este resultado é verdadeiro para algum $k > 0$. Logo, por indução,

$$\begin{aligned} M^{(k+1)}(M^{(1)}(A)) &= M^{(1)}(M^{(k)}(M^{(1)}(A))) \\ &= M^{(1)}(M^{(k+1)}(A)) \\ &= M^{(k+2)}(A). \end{aligned}$$

[P3] $E[M^{(k)}(A)] = M^{(k)}(E[A])$.

Quando $k = 0$, $E[M^{(0)}(A)] = E[A] = M^{(0)}(E[A])$. Considerando a hipótese de indução, temos

$$\begin{aligned} E[M^{(k+1)}(A)] &= E[M^{(1)}(M^{(k)}(A))] \\ &= E[\text{diag}\{HM^{(k)}(A)(H - 2I)\}] \\ &= \text{diag}\{HE[M^{(k)}(A)](H - 2I)\} \\ &= \text{diag}\{HM^{(k)}(E[A])(H - 2I)\} \\ &= M^{(1)}(M^{(k)}(E[A])) \\ &= M^{(k+1)}(E[A]). \end{aligned}$$

A idéia é utilizar a função M para calcular os vieses de $\widehat{\Omega}_i$ e $\widehat{\psi}_i$, a fim de construir uma seqüência recursiva de estimadores corrigidos. O índice i varia sempre de 0 a 3.

Lembre que o viés de um estimador $\widehat{\theta}$ para um parâmetro θ é igual ao valor esperado deste estimador menos o valor verdadeiro do parâmetro que queremos estimar. Assim, precisamos inicialmente calcular os valores esperados de $\widehat{\Omega}_i$ e $\widehat{\psi}_i$. Temos

$$\begin{aligned} E[\widehat{\Omega}_i] &= D_i E[\text{diag}\{\widehat{u}\widehat{u}'\}] \\ &= D_i \text{diag}\{(I - H)E[uu'](I - H)\} \\ &= D_i \text{diag}\{(I - H)\Omega(I - H)\} \\ &= D_i \text{diag}\{H\Omega(H - 2I) + \Omega\} \\ &= D_i\{M^{(1)}(\Omega) + \Omega\}. \end{aligned}$$

Utilizando a notação $B_{\widehat{\Omega}}(\Omega)$ para representar o viés do estimador $\widehat{\Omega}$ avaliado em Ω , temos

$$\begin{aligned} B_{\widehat{\Omega}_i}(\Omega) &= E[\widehat{\Omega}_i] - \Omega \\ &= D_i \text{diag}\{H\Omega(H - 2I)\} + D_i\Omega - \Omega. \end{aligned}$$

Usando a função recursiva $M^{(k)}$, obtemos

$$B_{\widehat{\Omega}_i}(\Omega) = D_i\{M^{(1)}(\Omega) + M^{(0)}(\Omega)\} - M^{(0)}(\Omega).$$

De forma semelhante, chegamos à expressão para o viés de estimadores da matriz de covariâncias do estimador de MQO do parâmetro β :

$$\begin{aligned} B_{\widehat{\psi}_i}(\psi) &= E[\widehat{\psi}_i] - \psi \\ &= P\{D_i(M^{(1)}(\Omega) + M^{(0)}(\Omega)) - M^{(0)}(\Omega)\}P'. \end{aligned}$$

A idéia é construir novos estimadores para Ω_i e ψ_i , que são definidos como os estimadores $\widehat{\Omega}_i$ e $\widehat{\psi}_i$ menos seus respectivos vieses estimados. Seja $B_{\widehat{\Omega}_i}(\widehat{\Omega}_i)$ o viés estimado do estimador $\widehat{\Omega}_i$ e defina

$$\begin{aligned} \widehat{\Omega}_i^{(1)} &= \widehat{\Omega}_i - B_{\widehat{\Omega}_i}(\widehat{\Omega}_i) \\ &= D_i\widehat{\Omega}_i - D_i(M^{(1)}(\widehat{\Omega}_i) + M^{(0)}(\widehat{\Omega}_i)) + M^{(0)}(\widehat{\Omega}_i) \\ &= \widehat{\Omega}_i - D_iM^{(1)}(\widehat{\Omega}_i). \end{aligned}$$

O próximo passo será construir $\widehat{\Omega}_i^{(2)}$ e para tanto necessitaremos calcular $B_{\widehat{\Omega}_i^{(1)}}(\widehat{\Omega}_i)$. Temos inicialmente que

$$B_{\widehat{\Omega}_i^{(1)}}(\Omega) = E[\widehat{\Omega}_i^{(1)}] - \Omega$$

$$\begin{aligned}
&= \mathbb{E}[\widehat{\Omega} - D_i M^{(1)}(\widehat{\Omega})] - \Omega \\
&= M^{(1)}(\Omega) + \Omega - \mathbb{E}[D_i M^{(1)}(\widehat{\Omega})] - \Omega \\
&= M^{(1)}(\Omega) - D_i M^{(1)} \mathbb{E}[\widehat{\Omega}] \\
&= M^{(1)}(\Omega) - D_i M^{(1)}(M^{(1)}(\Omega) + \Omega) \\
&= M^{(1)}(\Omega) - D_i M^{(2)}(\Omega) - D_i M^{(1)}(\Omega).
\end{aligned}$$

Agora podemos obter $\widehat{\Omega}_i^{(2)}$. Lembrando que precisamos avaliar a função viés em $\widehat{\Omega}$, temos

$$\begin{aligned}
\widehat{\Omega}_i^{(2)} &= \widehat{\Omega}_i^{(1)} - B_{\widehat{\Omega}_i^{(1)}}(\widehat{\Omega}) \\
&= \widehat{\Omega} - D_i M^{(1)}(\widehat{\Omega}) - \{M^{(1)}(\widehat{\Omega}) - D_i M^{(1)}(\widehat{\Omega}) - D_i M^{(2)}(\widehat{\Omega})\} \\
&= \widehat{\Omega} - M^{(1)}(\widehat{\Omega}) + D_i M^{(2)}(\widehat{\Omega}).
\end{aligned}$$

Já obtivemos os dois primeiros elementos da seqüência de estimadores corrigidos. Por indução matemática podemos demonstrar que a iteração k desta seqüência é dada por

$$\widehat{\Omega}_i^{(k)} = \sum_{j=0}^{k-1} [(-1)^j M^{(j)}(\widehat{\Omega})] + (-1)^k D_i M^{(k)}(\widehat{\Omega}), \quad k = 1, 2, \dots$$

Para $k = 1$ o resultado é trivial. Portanto, supondo que a relação acima é verdadeira para algum $k > 1$, vamos provar que vale também para $k + 1$. Faz-se necessário calcular

$$\begin{aligned}
B_{\widehat{\Omega}_i^{(k)}}(\Omega) &= \mathbb{E}[\widehat{\Omega}_i^{(k)}] - \Omega \\
&= \mathbb{E} \left[\sum_{j=0}^{k-1} (-1)^j M^{(j)}(\widehat{\Omega}) + (-1)^k D_i M^{(k)}(\widehat{\Omega}) \right] - \Omega \\
&= \sum_{j=0}^{k-1} (-1)^j \mathbb{E}[M^{(j)}(\widehat{\Omega})] + (-1)^k D_i \mathbb{E}[M^{(k)}(\widehat{\Omega})] - \Omega \\
&= \sum_{j=0}^{k-1} (-1)^j [\mathbb{E}[M^{(j)}(\widehat{\Omega}) - M^{(j)}(\Omega)] + M^{(j)}(\Omega)] \\
&\quad + (-1)^k D_i \mathbb{E}[M^{(k)}(\widehat{\Omega}) - M^{(k)}(\Omega)] + (-1)^k D_i M^{(k)}(\Omega) - \Omega \\
&= \sum_{j=0}^{k-1} (-1)^j [M^{(j+1)}(\Omega) + M^{(j)}(\Omega)] + (-1)^k D_i M^{(k+1)}(\Omega) + (-1)^k D_i M^{(k)}(\Omega) - \Omega,
\end{aligned}$$

dado que, $\mathbb{E}[M^{(j)}(\widehat{\Omega}) - M^{(j)}(\Omega)] = M^{(j+1)}(\Omega)$. Como este último somatório é uma soma telescópica,⁵ todos os termos se cancelam ficando apenas o primeiro termo mais ou menos o

⁵ Uma soma é dita ser telescópica quando a k -ésima parcela é cancelada pela $(k + 1)$ -ésima, para $k \geq 2$.

último termo. Desta forma,

$$\begin{aligned}
B_{\widehat{\Omega}_i^{(k)}}(\Omega) &= \Omega + (-1)^{k-1}M^{(k)}(\Omega) + (-1)^k D_i M^{(k+1)}(\Omega) + (-1)^k D_i M^{(k)}(\Omega) - \Omega \\
&= -(-1)^k M^{(k)}(\Omega) + (-1)^k D_i M^{(k+1)}(\Omega) + (-1)^k D_i M^{(k)}(\Omega) \\
&= (-1)^k \{D_i M^{(k+1)}(\Omega) + D_i M^{(k)}(\Omega) - M^{(k)}(\Omega)\}.
\end{aligned}$$

Temos assim

$$\begin{aligned}
\widehat{\Omega}_i^{(k+1)} &= \widehat{\Omega}_i^{(k)} - B_{\widehat{\Omega}_i^{(k)}}(\widehat{\Omega}) \\
&= \sum_{j=0}^{k-1} (-1)^j M^{(j)}(\widehat{\Omega}) + (-1)^k D_i M^{(k)}(\widehat{\Omega}) - (-1)^k \{D_i M^{(k+1)}(\widehat{\Omega}) \\
&\quad + D_i M^{(k)}(\widehat{\Omega}) - M^{(k)}(\widehat{\Omega})\} \\
&= \sum_{j=0}^{k-1} (-1)^j M^{(j)}(\widehat{\Omega}) + (-1)^{k+1} \{D_i M^{(k+1)}(\widehat{\Omega}) - M^{(k)}(\widehat{\Omega})\} \\
&= \sum_{j=0}^k (-1)^j M^{(j)}(\widehat{\Omega}) + (-1)^{k+1} D_i M^{(k+1)}(\widehat{\Omega}).
\end{aligned}$$

Desta forma, construímos uma seqüência de estimadores para Ω_i e, conseqüentemente, temos uma seqüência de estimadores corrigidos para a matriz de covariâncias ψ_i , que é dada por

$$\widehat{\psi}_i^{(k)} = P\widehat{\Omega}_i^{(k)}P', \quad (2.6)$$

onde $P = (X'X)^{-1}X'$. Uma expressão para o seu viés pode ser dada por

$$\begin{aligned}
B_{\widehat{\psi}_i^{(k)}}(\Omega) &= E(\widehat{\psi}_i^{(k)}) - \psi \\
&= PE(\widehat{\Omega}_i^{(k)})P' - P\Omega P' \\
&= P\{E(\widehat{\Omega}_i^{(k)}) - \Omega\}P' \\
&= P\{B_{\widehat{\Omega}_i^{(k)}}(\Omega)\}P' \\
&= (-1)^k P\{D_i M^{(k+1)}(\Omega) + D_i M^{(k)}(\Omega) - M^{(k)}(\Omega)\}P'.
\end{aligned}$$

2.4. Estatísticas Quase- t

O interesse aqui reside em testar a hipótese $\mathcal{H}_0 : \beta_k = \beta_k^{(0)}$ versus $\mathcal{H}_1 : \beta_k \neq \beta_k^{(0)}$ ou, de forma mais geral, $\mathcal{H}_0 : c'\beta = a$ versus $\mathcal{H}_1 : c'\beta \neq a$, onde c é um vetor $p \times 1$, a é uma

constante e β_k é o k -ésimo elemento do vetor de parâmetros β . Sob as Suposições [S1], [S2], [S3'], [S4] e [S6], a estatística

$$t = \frac{c'\hat{\beta} - a}{\sqrt{\widehat{\text{var}}(c'\hat{\beta})}} \quad (2.7)$$

tem distribuição t de Student com $n - p$ graus de liberdade. Na expressão (2.7) a estimativa da variância usada no denominador é a estimativa usual de mínimos quadrados ordinários. Na presença de heteroscedasticidade a estatística t não possui distribuição t de Student, continuando contudo a convergir em distribuição para uma $\mathcal{N}(0, 1)$ quando $n \rightarrow \infty$, pois $[\text{var}(c'\hat{\beta})]^{-\frac{1}{2}}(c'\hat{\beta} - a) \sim \mathcal{N}(0, 1)$ e $[\widehat{\text{var}}(c'\hat{\beta})]^{-\frac{1}{2}} \xrightarrow{p} [\text{var}(c'\hat{\beta})]^{-\frac{1}{2}}$. Logo,

$$\begin{aligned} t &= \frac{c'\hat{\beta} - a}{\sqrt{\widehat{\text{var}}(c'\hat{\beta})}} \frac{\sqrt{\text{var}(c'\hat{\beta})}}{\sqrt{\text{var}(c'\hat{\beta})}} \\ &= \frac{c'\hat{\beta} - a}{\sqrt{\text{var}(c'\hat{\beta})}} \frac{\sqrt{\text{var}(c'\hat{\beta})}}{\sqrt{\widehat{\text{var}}(c'\hat{\beta})}}. \end{aligned}$$

Uma vez que

$$\frac{\sqrt{\text{var}(c'\hat{\beta})}}{\sqrt{\widehat{\text{var}}(c'\hat{\beta})}} \xrightarrow{p} 1,$$

dado que o estimador da variância é consistente, temos que

$$\frac{c'\hat{\beta} - a}{\sqrt{\text{var}(c'\hat{\beta})}} \xrightarrow{d} \mathcal{N}(0, 1),$$

ou seja, $t \xrightarrow{d} \mathcal{N}(0, 1)$. Conseqüentemente, t^2 converge em distribuição para χ_1^2 . Se conseguirmos escrever o quadrado da estatística ‘quase- t ’, definida em (2.7), como uma razão de formas quadráticas de variáveis normais e considerarmos [S6], poderemos usar o algoritmo de Imhof (1961) e avaliar de forma exata a função de distribuição desta estatística.

O objetivo principal desta dissertação é avaliar numericamente a função de distribuição exata da estatística t^2 e compará-la com a distribuição qui-quadrado assintótica. Desta forma, iremos observar como os estimadores HC’s se comportam em variadas situações, tais como pequenas e grandes amostras, populações com diferentes tipos de variâncias, etc.

Cribari-Neto & Galvão (2002) construíram uma expressão em forma quadrática de variáveis normais para o denominador de t ; ver também Galvão (2000). Observe que

$$\phi = \text{var}(c'\hat{\beta}) = c'\psi_{\hat{\beta}}c.$$

Podemos definir uma seqüência de estimadores corrigidos para ϕ utilizando a equação (2.6) da seguinte forma:

$$\widehat{\phi}_i^{(k)} = c' \widehat{\psi}_i^{(k)} c = c' P \widehat{\Omega}_i^{(k)} P' c,$$

para $k = 1, 2, \dots$. Denotando $V_i = \text{diag}\{v_i v_i'\}$, com $v_i = D_i^{\frac{1}{2}} P' c$, temos

$$\begin{aligned} \widehat{\phi}_i^{(k)} &= c' P \left\{ \sum_{j=0}^{k-1} D_0^{\frac{1}{2}} (-1)^j M^{(j)}(\widehat{\Omega}) D_0^{\frac{1}{2}} + (-1)^k D_i^{\frac{1}{2}} M^{(k)}(\widehat{\Omega}) D_i^{\frac{1}{2}} \right\} P' c \\ &= c' P D_0^{\frac{1}{2}} \left\{ \sum_{j=0}^{k-1} (-1)^j M^{(j)}(\widehat{\Omega}) \right\} D_0^{\frac{1}{2}} P' c + c' P D_i^{\frac{1}{2}} \left\{ (-1)^k M^{(k)}(\widehat{\Omega}) \right\} D_i^{\frac{1}{2}} P' c \\ &= v_0' \left\{ \sum_{j=0}^{k-1} (-1)^j M^{(j)}(\widehat{\Omega}) \right\} v_0 + v_i' (-1)^k M^{(k)}(\widehat{\Omega}) v_i \\ &= \widehat{u}' \left\{ \sum_{j=0}^{k-1} (-1)^j M^{(j)}(V_0) + (-1)^k M^{(k)}(V_i) \right\} \widehat{u}. \end{aligned} \quad (2.8)$$

Para detalhes sobre a última passagem, ver Galvão (2000, pp. 24–25). Então,

$$\widehat{\phi}_i^{(k)} = \widehat{u}' Q_i^{(k)} \widehat{u},$$

onde $Q_i^{(k)} = \sum_{j=0}^{k-1} (-1)^j M^{(j)}(V_0) + (-1)^k M^{(k)}(V_i)$.

A fim de escrevermos a seqüência definida em (2.8) em forma quadrática de variáveis normais padrão, exigiremos que o vetor aleatório da forma quadrática tenha média zero e variância identidade. Para isto, utilizaremos um artifício matemático. Dado que $\widehat{u} = My$, onde $M = I - X(X'X)^{-1}X'$, então

$$\begin{aligned} \widehat{\phi}_i^{(k)} &= y' M Q_i^{(k)} M y \\ &= y' \Omega^{-\frac{1}{2}} \Omega^{\frac{1}{2}} M Q_i^{(k)} M \Omega^{\frac{1}{2}} \Omega^{-\frac{1}{2}} y \\ &= z' G_i^{(k)} z, \end{aligned}$$

onde $G_i^{(k)} = \Omega^{\frac{1}{2}} M Q_i^{(k)} M \Omega^{\frac{1}{2}}$ e $z = \Omega^{-\frac{1}{2}} y$. É fácil ver que o vetor z tem covariância unitária, pois

$$\begin{aligned} \text{cov}(z) &= \Omega^{-\frac{1}{2}} \text{cov}(y) \Omega^{-\frac{1}{2}} \\ &= \Omega^{-\frac{1}{2}} \text{E}[(y - \text{E}[y])(y - \text{E}[y])'] \Omega^{-\frac{1}{2}} \\ &= \Omega^{-\frac{1}{2}} \text{E}[(y - X\beta)(y - X\beta)'] \Omega^{-\frac{1}{2}} \\ &= \Omega^{-\frac{1}{2}} \text{E}[uu'] \Omega^{-\frac{1}{2}} \\ &= \Omega^{-\frac{1}{2}} \Omega \Omega^{-\frac{1}{2}} \\ &= I_n. \end{aligned}$$

Para que o vetor que define a forma quadrática de interesse tenha média zero é suficiente fazer uma translação de sua média. Assim sendo, definiremos

$$w = z - \theta,$$

onde $\theta = E[z]$. Observe que θ é ortogonal a $G_i^{(k)}$, pois

$$\theta' G_i^{(k)} = \beta' X' M Q_i^{(k)} M \Omega^{\frac{1}{2}} = 0,$$

uma vez que $X'M = X'(I - H) = 0$. Note ainda que

$$\begin{aligned} (z - \theta)' G_i^{(k)} (z - \theta) &= z' G_i^{(k)} z - 2\theta' G_i^{(k)} z + \theta' G_i^{(k)} \theta \\ &= z' G_i^{(k)} z. \end{aligned}$$

Como queríamos, escrevemos a seqüência corrigida $\widehat{\phi}_i^{(k)}$ como forma quadrática, de maneira tal que o vetor que define esta forma tem média zero e variância igual à matriz identidade, ou seja,

$$\widehat{\phi}_i^{(k)} = w' G_i^{(k)} w,$$

onde $w \sim \mathcal{N}(0, I_n)$ devido a [S6].

O próximo passo é escrever o numerador da estatística

$$t^2 = \frac{(c'\widehat{\beta} - a)^2}{\widehat{\text{var}}(c'\widehat{\beta})} \quad (2.9)$$

como forma quadrática em uma variável normal. Dado que $\widehat{\beta} = \beta + (X'X)^{-1}X'u$, onde $u \sim \mathcal{N}(0, \Omega)$, podemos escrever $\widehat{\beta} = \beta + (X'X)^{-1}X'\Omega^{\frac{1}{2}}w$, com w definido acima. Note que o numerador de t^2 em (2.9) é

$$\begin{aligned} (c'\widehat{\beta} - a)^2 &= \{c'\beta + c'(X'X)^{-1}X'\Omega^{\frac{1}{2}}w - a\}' \{c'\beta + c'(X'X)^{-1}X'\Omega^{\frac{1}{2}}w - a\} \\ &= \{(c'\beta - a) + c'(X'X)^{-1}X'\Omega^{\frac{1}{2}}w\}' \{(c'\beta - a) + c'(X'X)^{-1}X'\Omega^{\frac{1}{2}}w\} \\ &= (c'\beta - a)'(c'\beta - a) + 2(c'\beta - a)c'(X'X)^{-1}X'\Omega^{\frac{1}{2}}w \\ &\quad + w'\Omega^{\frac{1}{2}}X(X'X)^{-1}cc'(X'X)^{-1}X'\Omega^{\frac{1}{2}}w. \end{aligned}$$

Assim, temos

$$t^2 = \frac{w'Rw}{w'G_i^{(k)}w} + \frac{(c'\beta - a)'(c'\beta - a) + 2(c'\beta - a)c'(X'X)^{-1}X'\Omega^{\frac{1}{2}}w}{w'G_i^{(k)}w}, \quad (2.10)$$

onde $R = \Omega^{\frac{1}{2}}X(X'X)^{-1}cc'(X'X)^{-1}X'\Omega^{\frac{1}{2}}$, $G_i^{(k)} = \Omega^{\frac{1}{2}}MQ_i^{(k)}M\Omega^{\frac{1}{2}}$ e $w \sim \mathcal{N}(0, I_n)$. Sob a hipótese $\mathcal{H}_0 : c'\beta = a$, tal como foi definido no início da Seção 2.4, o segundo termo do lado

direito de (2.10) é zero e t^2 fica escrita em razão de formas quadráticas de variáveis normais padrão como

$$t^2 = \frac{w' R w}{w' G_i^{(k)} w}. \quad (2.11)$$

Portanto, dado $\gamma > 0$, temos

$$\begin{aligned} P(t^2 \leq \gamma \mid c'\beta = a) &= P_{\mathcal{H}_0}(t^2 \leq \gamma) \\ &= P_{\mathcal{H}_0}\left(\frac{w' R w}{w' G_i^{(k)} w} \leq \gamma\right). \end{aligned} \quad (2.12)$$

O próximo passo deste trabalho é avaliar numericamente esta probabilidade utilizando o algoritmo de Imhof (1961) e compará-la com $P(\chi_1^2 \leq \gamma)$ variando a constante γ , considerando os estimadores HC's em várias situações. O desempenho de t^2 é influenciado diretamente pelo comportamento do estimador $\widehat{\phi}_i^{(k)}$ utilizado em seu denominador. O viés deste estimador pode ser calculado pela expressão

$$\begin{aligned} B_{\widehat{\phi}_i^{(k)}}(\Omega) &= E[\widehat{\phi}_i^{(k)}] - \phi \\ &= E[w' G_i^{(k)} w] - c' \psi c \\ &= \text{tr}(G_i^{(k)}) - c' \psi c, \end{aligned}$$

dado que $E[w' G_i^{(k)} w] = \text{tr}(G_i^{(k)}) + E[w]' G_i^{(k)} E[w]$; ver Searle (1977, p. 55). Sua variância é dada por

$$\text{var}(\widehat{\phi}_i^{(k)}) = E\left[(w' G_i^{(k)} w)^2\right] - \left(E[w' G_i^{(k)} w]\right)^2.$$

Como $w \sim \mathcal{N}(0, I)$, um resultado de Seber (1977, p. 16) nos garante que

$$\text{var}(\widehat{\phi}_i^{(k)}) = 2\text{tr}\left([G_i^{(k)}]^2\right).$$

2.5. O Algoritmo de Imhof

Imhof (1961) desenvolveu métodos exatos para avaliar a função de distribuição de uma forma quadrática de variáveis normais. Seja $d \sim \mathcal{N}(0, \Sigma)$, onde d é um vetor $n \times 1$, 0 é um vetor $n \times 1$ de zeros e Σ é a matriz de covariâncias de d . Inicialmente n é assumido ser par. A forma quadrática

$$Q = d' A d,$$

onde A é uma matriz simétrica de ordem $n \times n$, pode ser escrita na forma

$$Q = \sum_{r=1}^m \lambda_r \chi_{h_r}^2,$$

onde os λ_r 's são os autovalores da matriz $A\Sigma$, os h_r 's são as multiplicidades dos respectivos autovalores, m é o número de autovalores distintos e $\chi_{h_r}^2$'s são variáveis independentes com distribuições qui-quadrado com h_r graus de liberdade; ver Scheffé (1959, p. 418).

Dada a forma quadrática Q definida acima, seja γ um escalar maior que zero, $\lambda_1 > \dots > \lambda_p > 0 > \lambda_{p+1} > \dots > \lambda_m$, $h_r = 2\nu_r$, para $r = 1, \dots, m$, com ν_r um inteiro positivo, $n = \sum_{r=1}^m \nu_r$ e $q = \sum_{r=1}^p \nu_r$. Imhof (1961), a partir dos resultados em Box (1954), demonstrou que

$$P(Q > \gamma) = \sum_{k=1}^p \frac{1}{(\nu_k - 1)!} \left[\frac{\partial^{\nu_k-1}}{\partial \lambda^{\nu_k-1}} F_k(\lambda, \gamma) \right]_{\lambda=\lambda_k}, \quad (2.13)$$

onde

$$F_k(\lambda, \gamma) = \lambda^{n-1} e^{-\frac{\gamma}{2\lambda}} \prod_{\substack{r=1 \\ r \neq k}}^m (\lambda - \lambda_r)^{-\nu_r}.$$

A probabilidade definida acima pode ser calculada através da expressão dada em Gil-Pelaez (1951), a saber:

$$P(Q > \gamma) = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \frac{\text{sen}\{\theta(u)\}}{u\rho(u)} du, \quad (2.14)$$

onde

$$\theta(u) = \frac{1}{2} \sum_{r=1}^m h_r \tan^{-1}(\lambda_r u) - \frac{1}{2} \gamma u \quad \text{e} \quad \rho(u) = \prod_{r=1}^m (1 + \lambda_r^2 u^2)^{\frac{1}{4} h_r}.$$

As expressões (2.13) e (2.14) podem ser generalizadas para o caso onde a variável aleatória d tem distribuição normal com vetor de médias diferente de zero e quando n é ímpar (Imhof, 1961, pp. 421–422). Para maiores detalhes sobre o algoritmo proposto por Imhof, ver Farebrother (1990).

A função `Imhof(x, A, B, m, S)`, definida abaixo e escrita na linguagem de programação `0x` (Doornik, 2001), avalia numericamente a probabilidade de um quociente de formas quadráticas de variáveis normais ser maior que um escalar $x > 0$, sendo A e B matrizes quadradas de ordem n , m o vetor de médias da variável normal e S sua matriz de covariâncias. Ou seja, esta função avalia

$$P\left(\frac{d'Ad}{d'Bd} > x\right),$$

onde $d \sim \mathcal{N}(m, S)$. O código da função em `0x` encontra-se apresentado abaixo.

CÓDIGO DA FUNÇÃO `Imhof(x, A, B, S)`

```
#include <oxstd.h>
#include <oxfloat.h>
#include <quadpack.h>

const decl pi = M_PI;
static decl l;
```

```

static decl d;
static decl c;
mgf(const u)
{
  decl eps = 0.5*(sumc(atan(1*u)+(d.^2).*1*u./(1+((1*u).^2)))-c*u);
  decl gam = prodc(((1+((1*u).^2)).^0.25).*exp(0.5*((d.*1*u).^2)/(1+(1*u).^2)));
  return (sin(eps)/(u*gam));
}

Imhof(const x, const A, const B, const m, const S)
{
  decl Q, V, result, abserr;
  decl P = choleski(S);
  if (B == 0)
  {
    Q = P'((A+A')/2)*P;
    eigensym(Q,&l,&V);
    c = x;
  }
  else
  {
    Q = A - B*x;
    Q = (P'((Q+Q')/2)*P);
    eigensym(Q,&l,&V);
    c = 0;
  }
  d = V'invert(P)*m;
  l = l';
  d = selectifr(d,l);
  l = selectifr(l,l);
  QAGI(mgf,0,1,&result,&abserr);
  return (0.5 - result/pi);
}

```

As funções `Ox` acima foram escritas por Peter Boswijk (University of Amsterdam) e obtidas diretamente do Professor Jurgen Doornik (Nuffield College, University of Oxford). (Para implementações em `Pascal` do algoritmo de Imhof, ver Farebrother, 1990.)

No caso onde a variável d tem vetor de médias zero, o cálculo desta probabilidade é reduzido à expressão (2.14). A probabilidade que queremos avaliar em (2.12) é um quociente de formas quadráticas em variáveis normais com média zero. O código acima faz a seguinte transformação:

$$P\left(\frac{d'Ad}{d'Bd} > c\right) = P(d'Ad - d'Bdc > 0) = P(d'(A - cB)d > 0),$$

onde A e B encontram-se definidas acima.

A função `mgf(u)`, definida na linha 8 do código acima, no caso particular de $d \sim \mathcal{N}(0, \Sigma)$, calcula a expressão do integrando em (2.14) e `QAGI(mgf, 0, 1, &result, &abserr)` na penúltima linha do código calcula a integral em (2.14).

Capítulo 3

Análise Exata de Testes por Integração Numérica

3.1. Introdução

O objetivo do presente capítulo é usar o método de Imhof aplicado a estatísticas quase- t construídas com base nos estimadores consistentes de variâncias discutidos anteriormente. Chesher & Austin (1991) realizaram análise semelhante, mas não consideraram estimadores de variâncias que incorporam ajuste de viés. No presente capítulo consideraremos tanto estimadores não-corrigidos quanto estimadores corrigidos, generalizando assim a análise realizada por estes autores. Mais especificamente, o nosso objetivo será avaliar se a utilização de estimadores de variâncias corrigidos por viés na construção de estatísticas de teste quase- t aproxima bem a distribuição nula destas estatísticas da distribuição assintótica de referência. Isto é relevante uma vez que as estimativas de variâncias são comumente utilizadas para realizar inferências sobre os parâmetros de regressão e não há garantia alguma de que estimadores com menor viés forneçam estatísticas de teste cujas distribuições nulas exatas (desconhecidas) se encontrem mais próximas da distribuição assintótica utilizada na realização do teste. Para tanto, nos restringiremos ao caso onde a variável resposta é normalmente distribuída. Os resultados apresentados a seguir mostram que nem sempre a utilização de estimadores melhorados na construção de estatísticas de teste conduz a inferências mais precisas.

3.2. Modelo e Teste

O modelo considerado para análise no presente capítulo é

$$y_t = \beta_0 + \beta_1 x_t + u_t, \quad t = 1, \dots, n,$$

onde y_t é a variável resposta, x_t é a variável independente e u_t é o erro do modelo. Este erro possui média zero e variância $\sigma_t^2 = \exp(\alpha_1 x_t + \alpha_2 x_t^2)$. Definimos a seguinte medida do grau de heteroscedasticidade:

$$\lambda = \frac{\max\{\sigma_t^2\}}{\min\{\sigma_t^2\}}.$$

Note que: (i) sob homoscedasticidade, $\lambda = 1$; (ii) quando as variâncias dos erros não são constantes, quanto maior λ , maior o grau de heteroscedasticidade. De acordo com Chesher & Jewitt (1987), heteroscedasticidade moderada ocorre quando $1 < \lambda < 2$. Ao longo do

capítulo será assumido que u_t possui distribuição normal, o que permitirá a utilização do algoritmo de Imhof.

O nosso interesse recairá sobre o teste de hipótese nula $\mathcal{H}_0 : c'\beta = a$, onde c é um dado vetor $p \times 1$ e a é um determinado escalar. No decorrer da análise usaremos $c = (0, 1)'$ e $a = 1$, o que corresponde ao teste de $\mathcal{H}_0 : \beta_1 = 1$. A hipótese alternativa é bicaudal, isto é, $\mathcal{H}_1 : \beta_1 \neq 1$. A estatística quase- t utilizada é da forma (2.7) e possui distribuição nula limite $\mathcal{N}(0, 1)$. Trabalharemos com t^2 como dado em (2.9), que, sob a hipótese nula \mathcal{H}_0 , possui distribuição assintótica χ_q^2 , onde q denota o número de restrições sob teste. No caso de interesse, $q = 1$ e

$$t^2 = \frac{(\widehat{\beta}_1 - 1)^2}{\widehat{\text{var}}(\widehat{\beta}_1)}. \quad (3.1)$$

Diferentes estimadores consistentes da variância de $\widehat{\beta}_1$ serão usados no denominador da estatística de teste dada em (3.1). Como visto na Seção 2.4, esta estatística pode ser escrita sob \mathcal{H}_0 como uma razão de formas quadráticas de vetores normais padrão; ver equação (2.11).

3.3. Resultados Numéricos

A avaliação será realizada comparando a distribuição exata, sob \mathcal{H}_0 , da estatística dada em (3.1) com a distribuição χ_1^2 , que corresponde à distribuição limite de (3.1) e que é tipicamente usada para se obter um valor crítico aproximado para o teste. O cálculo da distribuição exata de t^2 será realizado usando o algoritmo de Imhof (1961). Para tanto, nós usaremos o resultado derivado no Capítulo 2; ver (2.12).

No que diz respeito ao modelo usado, a avaliação será feita com base no modelo linear de regressão descrito acima com $\beta_0 = \beta_1 = 1$. Inicialmente consideraremos o caso em que $n = 25$. Os valores de x para este tamanho amostral serão obtidos como números aleatórios independentes de uma distribuição uniforme $\mathcal{U}(0, 1)$ e replicados duas vezes para obtermos amostras de 50 observações. Neste caso utilizaremos $\alpha_1 = \alpha_2 = 0.000$ e $\alpha_1 = \alpha_2 = 3.100$, que conduzem aos seguintes valores de λ : 1.000 (homoscedasticidade) e 101.661 (heteroscedasticidade forte). Este caso será referido como “desenho 1”. O “desenho 2” será similar, mas aqui os valores de x serão obtidos independentemente de uma distribuição t_3 . Neste caso, $\alpha_1 = \alpha_2 = 0.000$ e $\alpha_1 = \alpha_2 = 0.415$, conduzindo a $\lambda = 1.000$ (homoscedasticidade) e $\lambda = 101.513$ (heteroscedasticidade forte). O segundo desenho objetiva estudar o efeito de pontos de alavanca na qualidade da inferência realizada.

A presença de pontos de alavanca no segundo desenho numérico pode ser identificada observando-se os valores de h_t , $t = 1, \dots, n$, os elementos diagonais da matriz H . Note que $\sum_{t=1}^n h_t = \text{tr}(H) = \text{tr}[X(X'X)^{-1}X'] = p$; assim, os h_t 's têm valor médio igual a p/n . Uma regra comumente utilizada é considerar como potencialmente influentes as observações

associadas a valores de h_t que excedam duas ou três vezes a média, ou seja, que excedam $2p/n$ ou $3p/n$ (Judge et al., 1988, p. 893). A Tabela 3.1 apresenta os valores máximo e mínimo de h_t para os dois cenários analisados, comparando-os com as quantidades $2p/n$ e $3p/n$. Quando os valores da covariada são obtidos de uma distribuição t_3 , há pontos de alavanca claros; para três observações, $h_t > 2p/n$ para $t = 1, 7, 11$ ($n = 25$).

Tabela 3.1. Análise de observações influentes.

n	$x \sim U(0, 1)$		$x \sim t_3$		$2p/n$	$3p/n$
	h_{min}	h_{max}	h_{min}	h_{max}		
25	0.040	0.143	0.041	0.245	0.160	0.240
50	0.020	0.072	0.020	0.123	0.080	0.120

Os gráficos apresentados a seguir contêm duas curvas, a saber: a distribuição nula assintótica da estatística de teste (χ_1^2), traçada com linha sólida, e as distribuições exatas calculadas, apresentadas visualmente com linhas tracejadas. Os testes avaliados utilizam os estimadores HC0, HC1, HC2 e HC3, e também suas respectivas versões corrigidas por viés, baseadas em uma, duas e até cinco iterações. Os quantis assintóticos de nível 0.90 e 0.95, a título de referência, são 2.706 e 3.841.

A Figura 3.1 apresenta o caso onde a covariada é obtida de uma distribuição uniforme, o tamanho amostral é 25, há forte heteroscedasticidade ($\lambda \approx 100$) e os estimadores usados no denominador da estatística de teste não incorporam correção de viés. Os quatro painéis, iniciando no canto superior esquerdo e terminando no canto inferior direito, correspondem às análises comparativas da distribuição nula limite e das distribuições nulas exatas das estatísticas de teste que usam os estimadores consistentes HC0 a HC3, aqui usados sem correção de viés. Inicialmente, notamos que as curvas tracejadas se encontram abaixo da curva sólida que representa a função de distribuição de uma variável aleatória χ_1^2 , inclusive em seu quantil de 0.95 (valor crítico assintótico de nível nominal 0.05), o que significa que os quatro testes sob consideração apresentam tendência liberal, ou seja, tendem a apresentar probabilidades de erro do tipo I superiores ao nível nominal selecionado. Quanto maior a distância entre as duas curvas no quantil assintótico de referência, maior a tendência do teste à super-rejeição. (No decorrer da análise consideraremos a situação onde este nível nominal é igual a 0.05.) Por exemplo, as funções de distribuição exatas das quatro estatísticas de teste, que utilizam os estimadores HC0, HC1, HC2 e HC3, no ponto 3.841 (quantil assintótico de 95%), são, respectivamente, iguais a 0.8811, 0.8938, 0.9007 e 0.9181. Notamos que o pior teste é aquele cuja estatística é construída usando o estimador HC0 (estimador de White), seguido dos casos onde os estimadores usados são HC1, HC2 e HC3. Assim, o teste cuja estatística utiliza o estimador HC3 é o que tenderá a apresentar menor distorção de tamanho.

A Figura 3.2 apresenta gráficos semelhantes, mas desta vez para testes cujas estatísticas

utilizam estimativas de variância corrigidas por viés com base em uma iteração do mecanismo seqüencial de redução de viés. Ao compararmos os gráficos apresentados nesta figura com aqueles dados na Figura 3.1, notamos que a utilização de estimadores corrigidos por viés melhora em alguns casos a aproximação assintótica do teste quase- t . Isto é claro principalmente para o teste cuja estatística utiliza o estimador de White (HC0): a curva tracejada encontra-se mais próxima da curva sólida no painel superior esquerdo da Figura 3.2 do que no correspondente painel da Figura 3.1. Nota-se ainda alguma melhoria para o teste cuja estatística utiliza o estimador HC1 e uma melhoria muito reduzida para o teste cuja estatística é construída com base no estimador HC2. É interessante notar, contudo, que a utilização de estimadores do tipo HC3 corrigidos por viés na construção da estatística de teste não melhora a aproximação assintótica dada pela distribuição nula assintótica de referência; ao contrário, esta aproximação é deteriorada. As funções de distribuição exatas das quatro estatísticas de teste avaliadas no ponto 3.841 são, respectivamente, 0.9020, 0.9017, 0.9026 e 0.9053, respectivamente.

As Figuras 3.3 e 3.4 apresentam gráficos similares aos fornecidos na Figura 3.2, mas agora usando duas e cinco iterações, respectivamente, do esquema iterativo de ajuste de viés. Notamos que o teste que mais se beneficia do aumento do número de iterações de correção de viés para obtenção da estimativa da variância usada na construção da estatística de teste é aquele cuja estatística se baseia no estimador HC0. A deterioração da qualidade da aproximação assintótica do teste que usa o estimador do tipo HC3 é mais uma vez verificada.

O efeito do aumento do tamanho amostral sobre os gráficos apresentados pode ser constatado analisando-se as Figuras 3.5 a 3.8, que correspondem aos casos descritos nas Figuras 3.1 a 3.4, mas agora utilizando 50 observações; ou seja, o tamanho amostral é dobrado. Este aumento do tamanho amostral é feito, como explicado anteriormente, replicando-se cada valor da covariada. Uma inspeção visual das Figuras 3.5 a 3.8 revela que, como esperado, a qualidade da aproximação assintótica dos testes quase- t melhora quando o número de observações cresce. As conclusões obtidas anteriormente de que o teste que mais se beneficia do esquema de correção de viés da variância usada em sua estatística de teste é o que usa o estimador HC0 da variância e que a utilização de estimadores HC3 corrigidos por viés conduz a uma deterioração na aproximação assintótica de primeira ordem do teste se mantêm aqui, quando $n = 50$.

Uma outra conclusão que pode ser obtida das Figuras 3.1 a 3.8 é que a estatística de teste que emprega o estimador HC3 sem nenhuma correção de viés possui melhor aproximação assintótica pela distribuição χ^2 do que a que utiliza o estimador HC0 corrigido, mesmo utilizando-se várias iterações do esquema iterativo de correção de viés. Ou seja, no que se refere a inferências via testes quase- t o estimador HC3 sem correção de viés é o preferido, no sentido de que é o estimador que produz estatísticas de teste com melhor aproximação assintótica de primeira ordem.

O nosso próximo interesse reside em investigar o comportamento dos testes considerados em situações onde há pontos de avalanche na estrutura de regressão. As Figuras 3.9 a 3.12 consideram o caso onde pontos de avalanche são introduzidos ao se selecionar aleatória e independentemente os 25 valores da covariável de uma distribuição t_3 . Neste caso, temos que $\alpha_1 = \alpha_2 = 0.415$, o que resulta em $\lambda = 101.513$ (heteroscedasticidade forte). Os estimadores utilizados na construção das estatísticas de teste na Figura 3.9 não incorporam correção de viés. Notamos que neste caso o teste que possui melhor aproximação assintótica de primeira ordem é aquele cuja estatística de teste utiliza o estimador HC3, seguido pelos testes cujas estatísticas baseiam-se nos estimadores HC2, HC1 e HC0. Mais uma vez o estimador HC3 foi o preferido, ao passo que o estimador HC0 foi o que conduziu a testes associados com pior aproximação assintótica.

Ao compararmos a Figura 3.9 ($x \sim t_3$) com a Figura 3.1 ($x \sim \mathcal{U}(0, 1)$), notamos que a introdução de pontos de avalanche nitidamente deteriora a aproximação de primeira ordem na qual o teste quase- t se baseia. Para constatar este fato basta notar que as distâncias entre as linhas tracejadas e sólidas são amplificadas quando pontos de avalanche são introduzidos na estrutura de regressão (Figura 3.9). As funções de distribuição exatas das quatro estatísticas de teste avaliadas no ponto 3.841 são, respectivamente, 0.8400, 0.8573, 0.8826 e 0.9176. Notamos que o teste baseado no estimador HC3 foi o menos influenciado pela introdução de pontos de avalanche.

As Figuras 3.10 a 3.12 consideram testes cujas estatísticas utilizam estimadores corrigidos por viés, sendo estes baseados em esquemas iterativos com uma, duas e cinco iterações, respectivamente, sob forte heteroscedasticidade, $n = 25$ e covariável com pontos potencialmente influentes. Ao analisarmos seqüencialmente estas figuras, juntamente com a Figura 3.9 descrita acima, notamos que o teste que mais se beneficia do esquema de correção de viés que obtém estimativas mais precisas de variância usadas na estatística de teste é aquele que utiliza o estimador HC0. Mais uma vez também a utilização de estimativas do tipo HC3 corrigidas por viés deteriora, ao invés de melhorar, a aproximação assintótica de primeira ordem do teste. A conclusão obtida anteriormente de que o teste baseado no estimador HC3 sem correção de viés é preferível aos demais testes que usam estimadores corrigidos se mantém.

As Figuras 3.13 a 3.16 também correspondem ao caso de forte heteroscedasticidade e covariada com pontos potencialmente influentes, mas agora consideram a situação onde o tamanho amostral é duplicado, isto é, temos agora 50 observações. Como esperado, a qualidade da aproximação assintótica a todos os testes melhora à medida que o número de observações aumenta. As conclusões obtidas e listadas anteriormente também são válidas nesta situação.

A análise acima foi realizada com base em uma estrutura heteroscedástica de regressão. É importante também, contudo, investigar o comportamento dos diferentes testes quando

o modelo é homoscedástico, ou seja, quando as variâncias dos diferentes erros são de fato constantes. As Figuras 3.17 a 3.28 consideram modelos com erros homoscedásticos, variando contudo outros aspectos dos modelos e testes (pro exemplo, número de observações, inclusão ou não de pontos de alavanca, etc.). Quando $n = 25$ e $x \sim t_3$, chegamos a considerar dez passos do mecanismo iterativo de correção de viés, avançando assim ainda mais na seqüência de correção de viés. As conclusões qualitativas enumeradas para o caso heteroscedástico permanecem válidas quando as variâncias dos erros são iguais.

3.4. Conclusões

Em linhas gerais, a análise numérica exata realizada no presente capítulo conduz a importantes conclusões sobre testes quase- t em modelos lineares de regressão com possível heteroscedasticidade de forma desconhecida, a saber:

- (i) O teste que mais se beneficia de usar estatísticas de teste construídas a partir de estimativas de variâncias que incorporam correção de viés é aquele que utiliza o estimador HC0 (estimador de White).
- (ii) O teste com melhor aproximação assintótica de primeira ordem é aquele cuja estatística de teste é construída utilizando o estimador HC3 *sem correção de viés*.
- (iii) No caso do teste que utiliza o estimador HC3, o esquema de correção de viés do estimador da variância usado na definição da estatística quase- t deteriora, ao invés de melhorar, a aproximação assintótica de primeira ordem.
- (iv) O desempenho dos diferentes testes em amostras de tamanho pequeno a moderado é influenciado pela presença de pontos de alavanca na estrutura de regressão. O teste menos afetado por pontos de alavanca é o que utiliza o estimador HC3.

Os resultados contidos no presente capítulo tomados conjuntamente revelam que a estratégia preferencial de inferência em modelos normais lineares de regressão com heteroscedasticidade de forma desconhecida é a utilização do estimador consistente de variâncias do tipo HC3 para a construção da estatística quase- t . O esquema de correção de viés proposto por Cribari-Neto & Galvão (2002) conduz a testes mais precisos quando utilizado para produzir estimativas de variâncias a serem usadas por estatísticas de teste, exceto quando o ajuste de viés é aplicado ao estimador HC3. Neste caso, a correção de viés conduz a testes associados menos precisos, contrariamente ao que ocorre quando o esquema de redução de viés é aplicado a estimadores alternativos.

Figura 3.1. Distribuições exata e assintótica de t^2 , $n = 25$, $\alpha_1 = \alpha_2 = 3.1$, $\lambda = 101.661$, $x \sim \mathcal{U}(0, 1)$, sem correção.

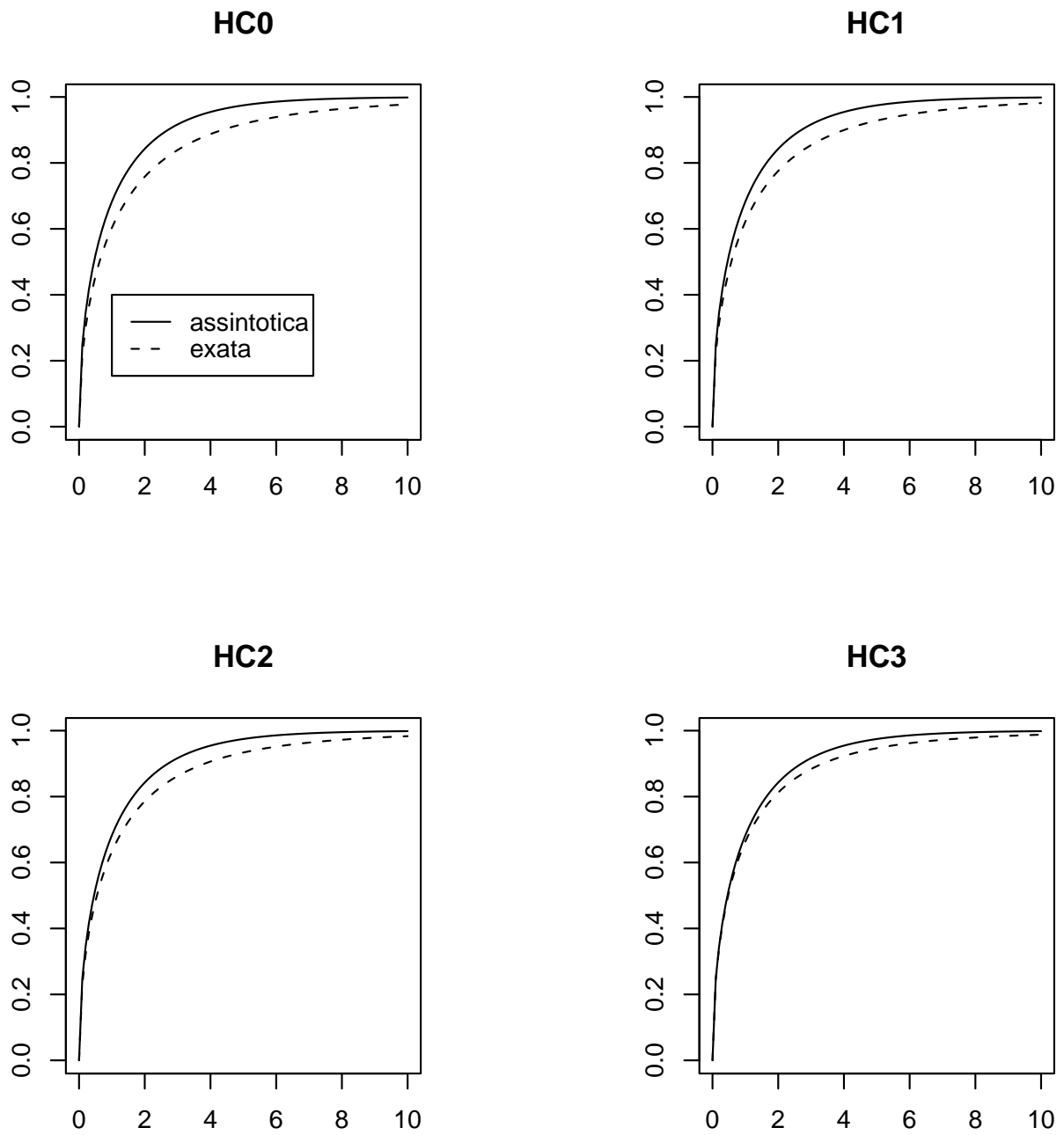


Figura 3.2. Distribuições exata e assintótica de t^2 , $n = 25$, $\alpha_1 = \alpha_2 = 3.1$, $\lambda = 101.661$, $x \sim \mathcal{U}(0, 1)$, correção com uma iteração.

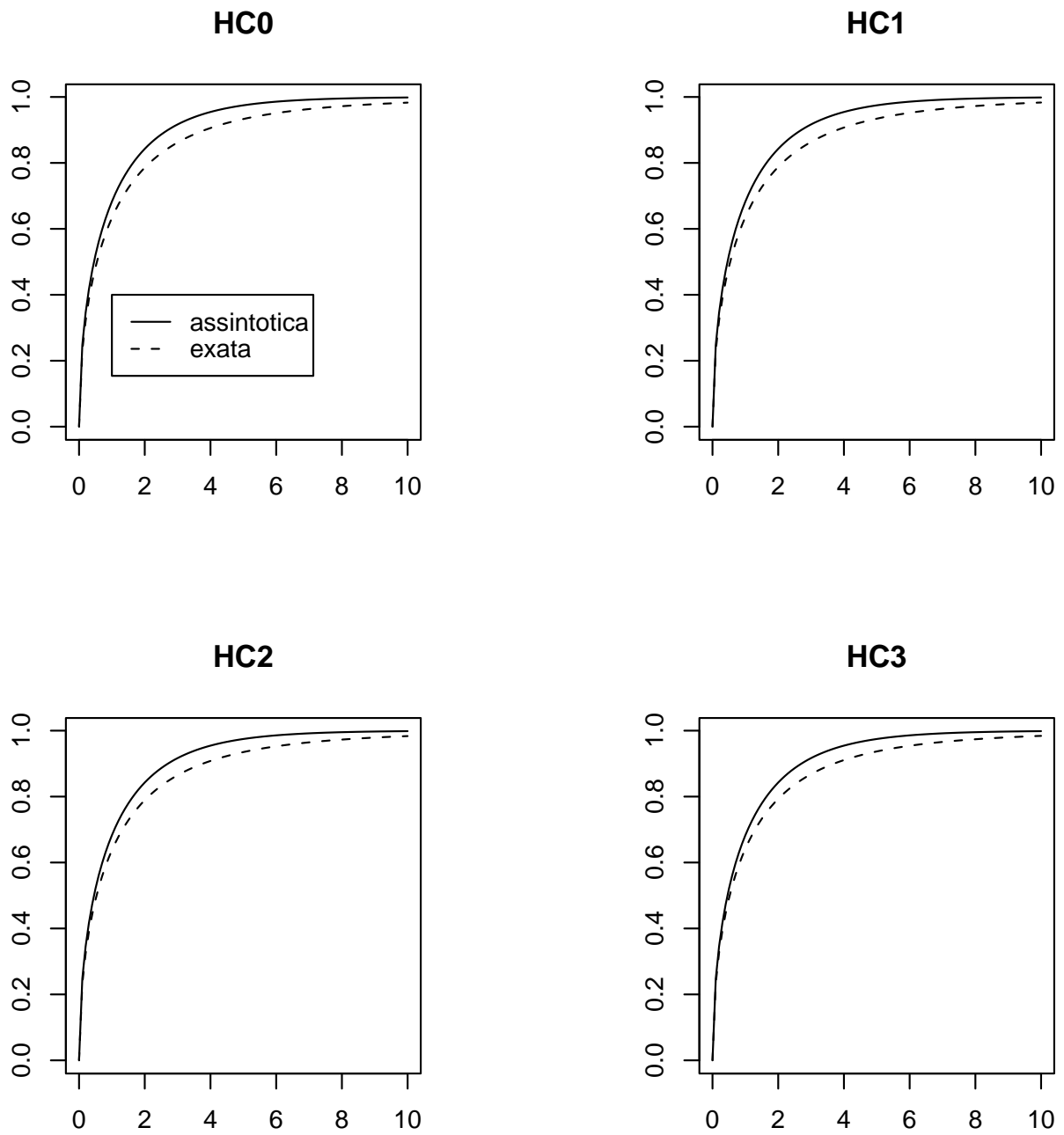


Figura 3.3. Distribuições exata e assintótica de t^2 , $n = 25$, $\alpha_1 = \alpha_2 = 3.1$, $\lambda = 101.661$, $x \sim \mathcal{U}(0, 1)$, correção com duas iterações.

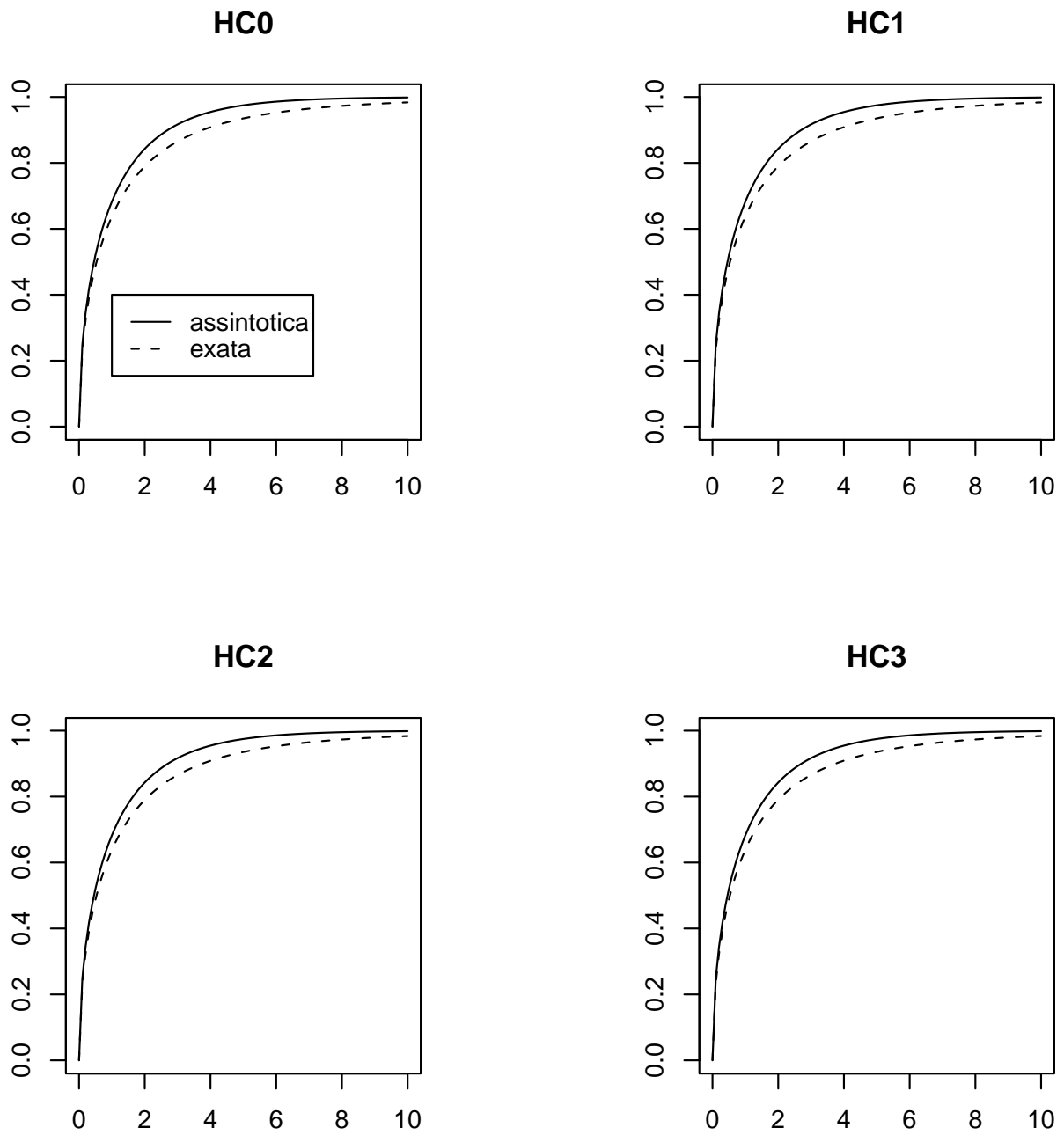


Figura 3.4. Distribuições exata e assintótica de t^2 , $n = 25$, $\alpha_1 = \alpha_2 = 3.1$, $\lambda = 101.661$, $x \sim \mathcal{U}(0, 1)$, correção com cinco iterações.

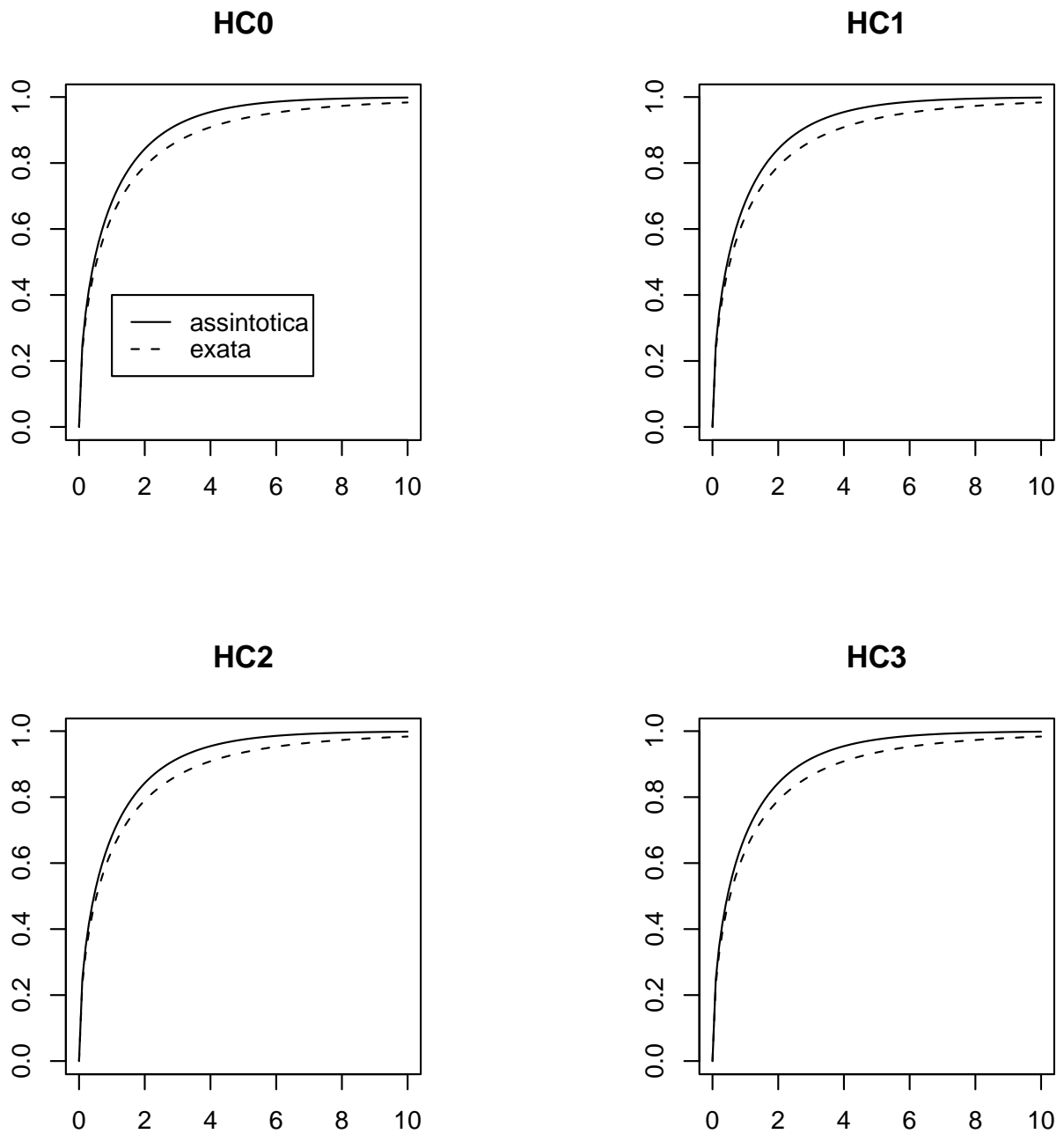


Figura 3.5. Distribuições exata e assintótica de t^2 , $n = 50$, $\alpha_1 = \alpha_2 = 3.1$, $\lambda = 101.661$, $x \sim \mathcal{U}(0, 1)$, sem correção.

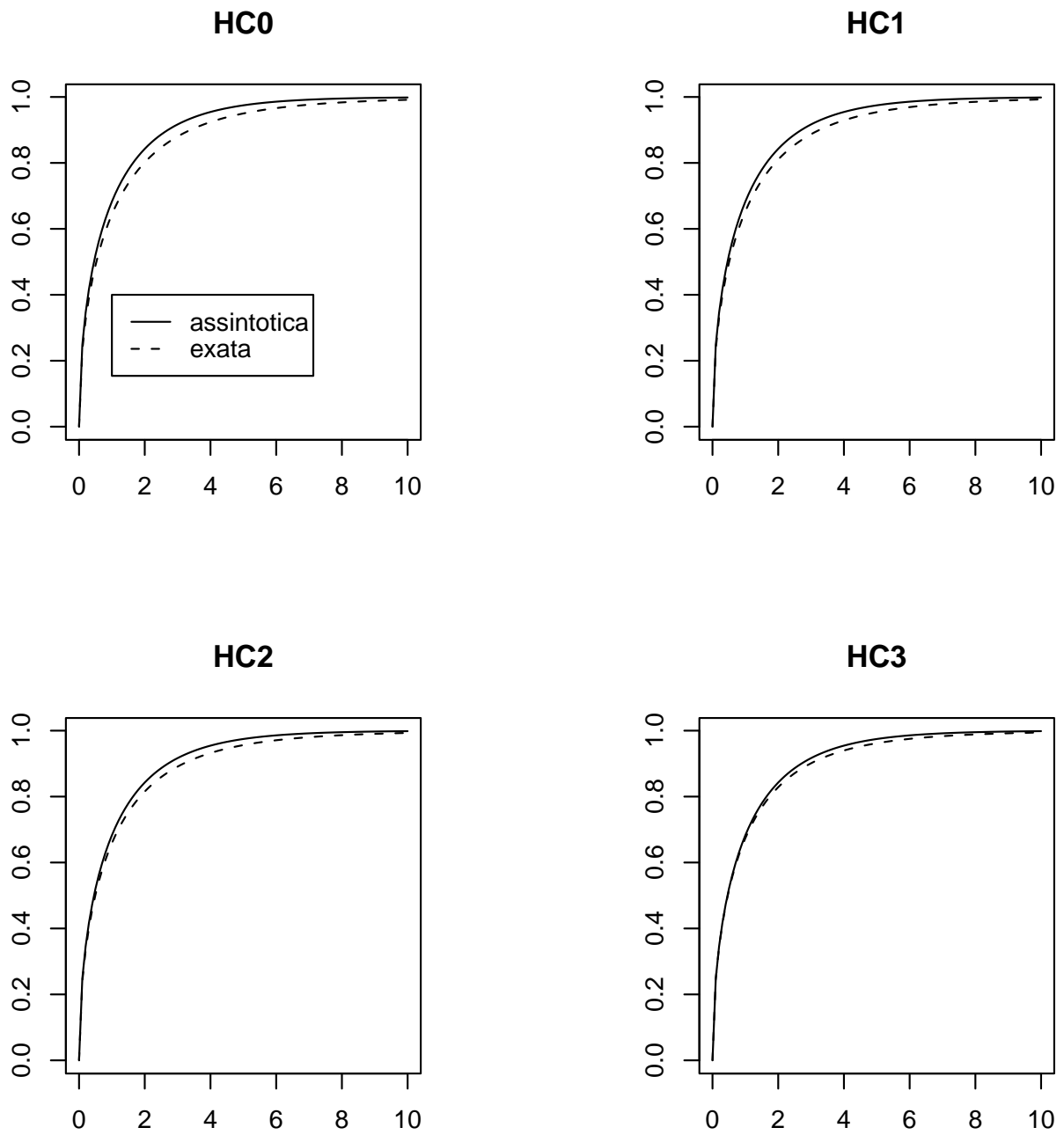


Figura 3.6. Distribuições exata e assintótica de t^2 , $n = 50$, $\alpha_1 = \alpha_2 = 3.1$, $\lambda = 101.661$, $x \sim \mathcal{U}(0, 1)$, correção com uma iteração.

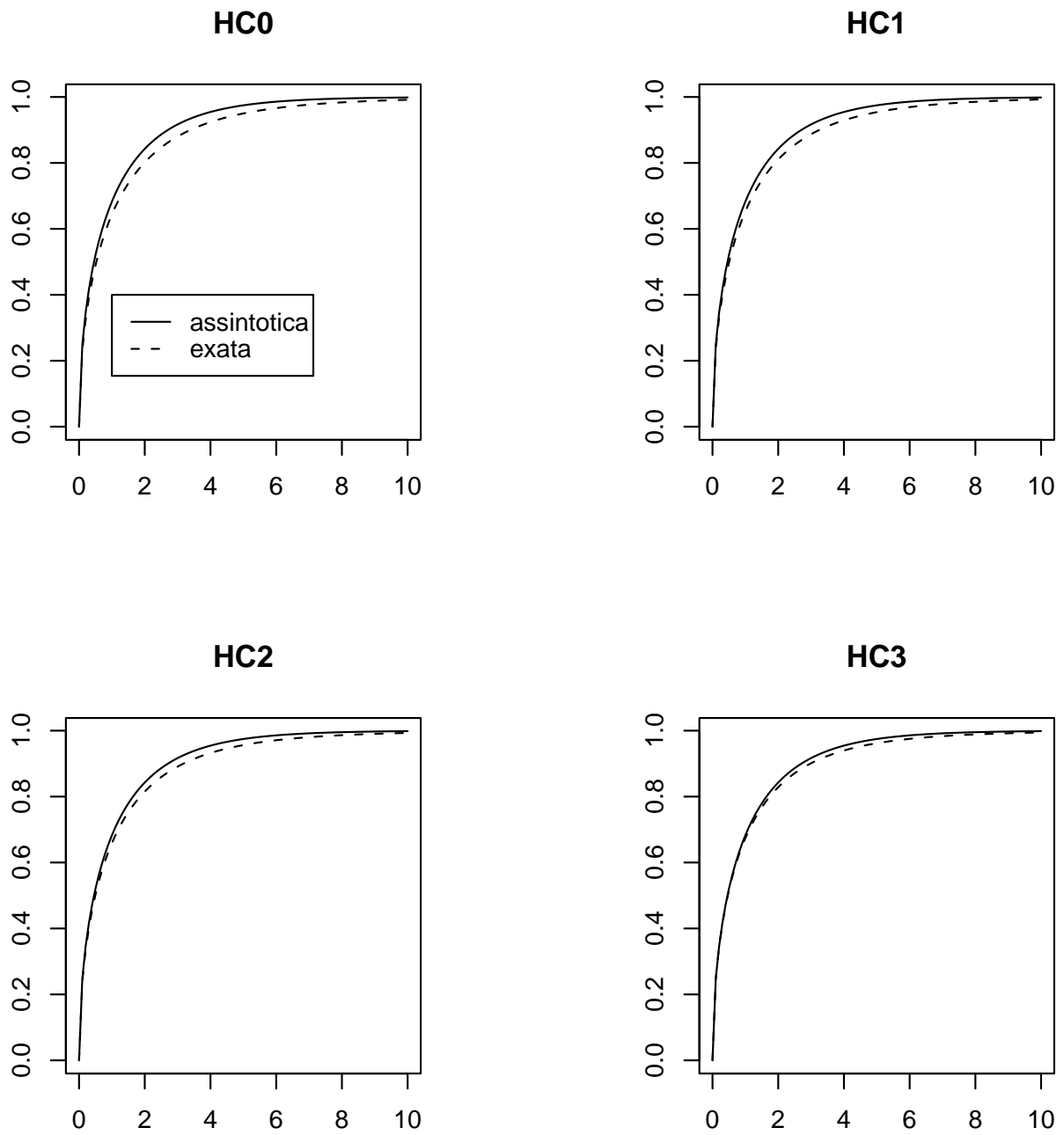


Figura 3.7. Distribuições exata e assintótica de t^2 , $n = 50$, $\alpha_1 = \alpha_2 = 3.1$, $\lambda = 101.661$, $x \sim \mathcal{U}(0, 1)$, correção com duas iterações.

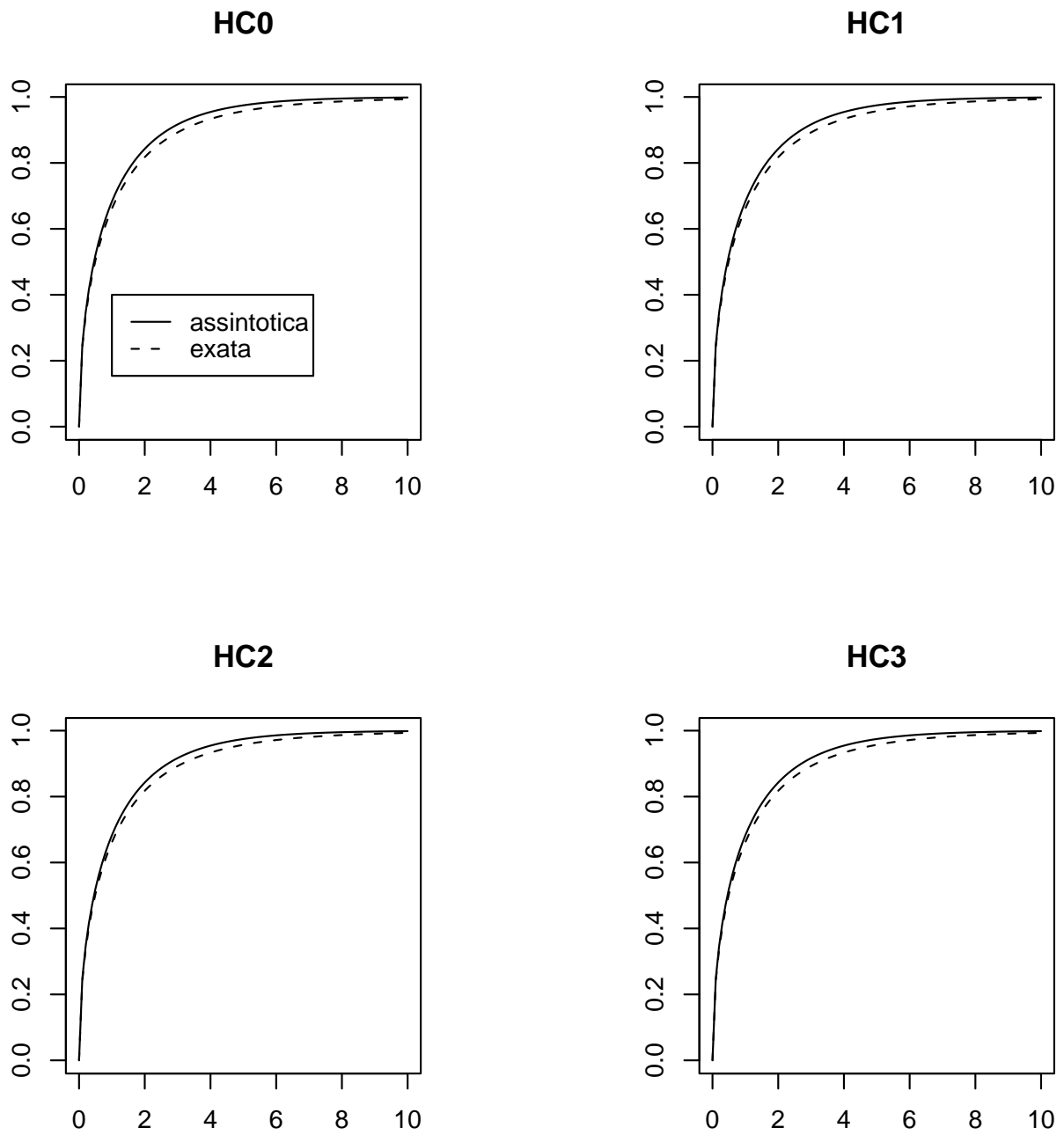


Figura 3.8. Distribuições exata e assintótica de t^2 , $n = 50$, $\alpha_1 = \alpha_2 = 3.1$, $\lambda = 101.661$, $x \sim \mathcal{U}(0, 1)$, correção com cinco iterações.

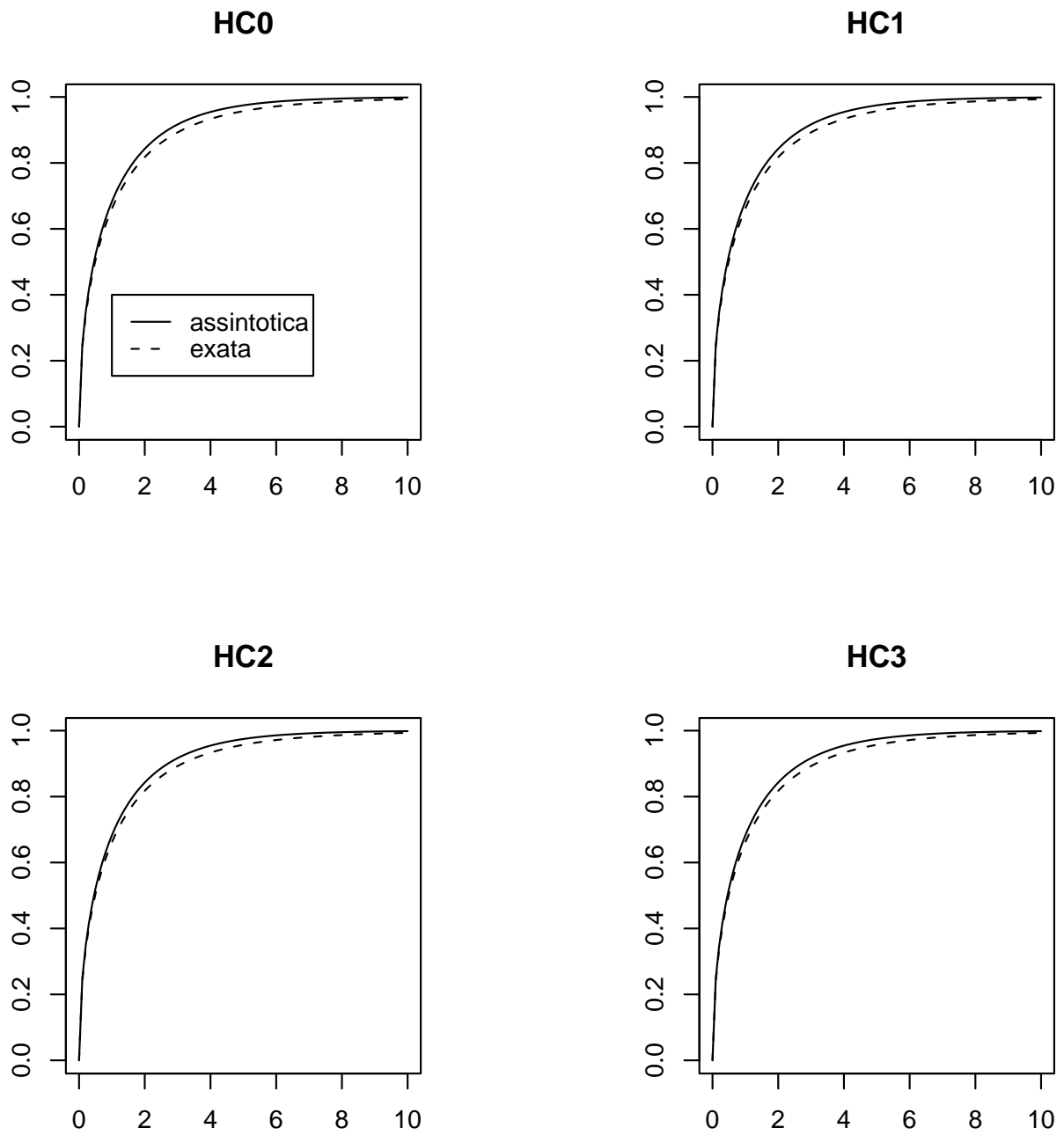


Figura 3.9. Distribuições exata e assintótica de t^2 , $n = 25$, $\alpha_1 = \alpha_2 = 0.415$, $\lambda = 101.513$, $x \sim t_3$, sem correção.

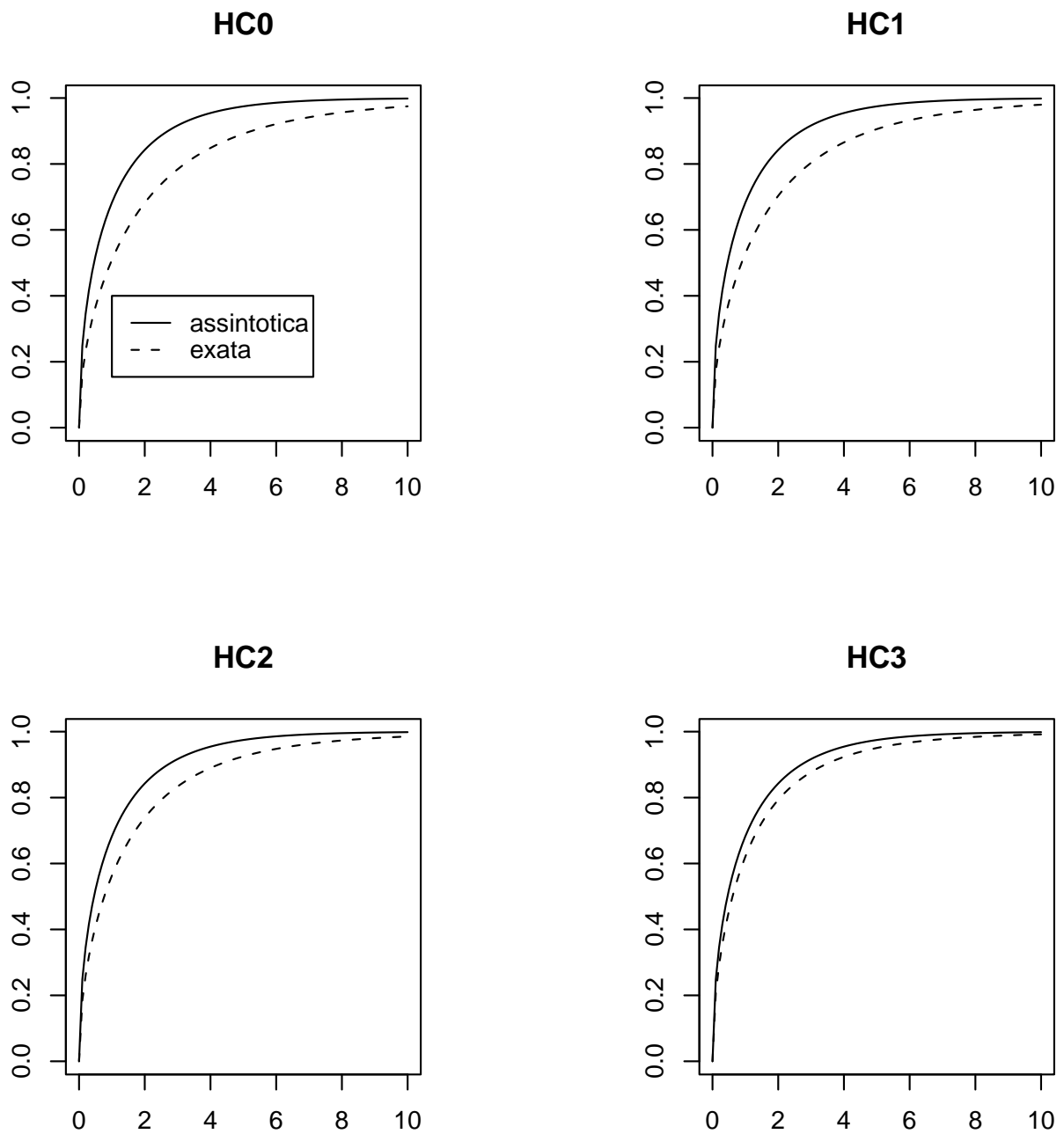


Figura 3.10. Distribuições exata e assintótica de t^2 , $n = 25$, $\alpha_1 = \alpha_2 = 0.415$, $\lambda = 101.513$, $x \sim t_3$, correção com uma iteração.

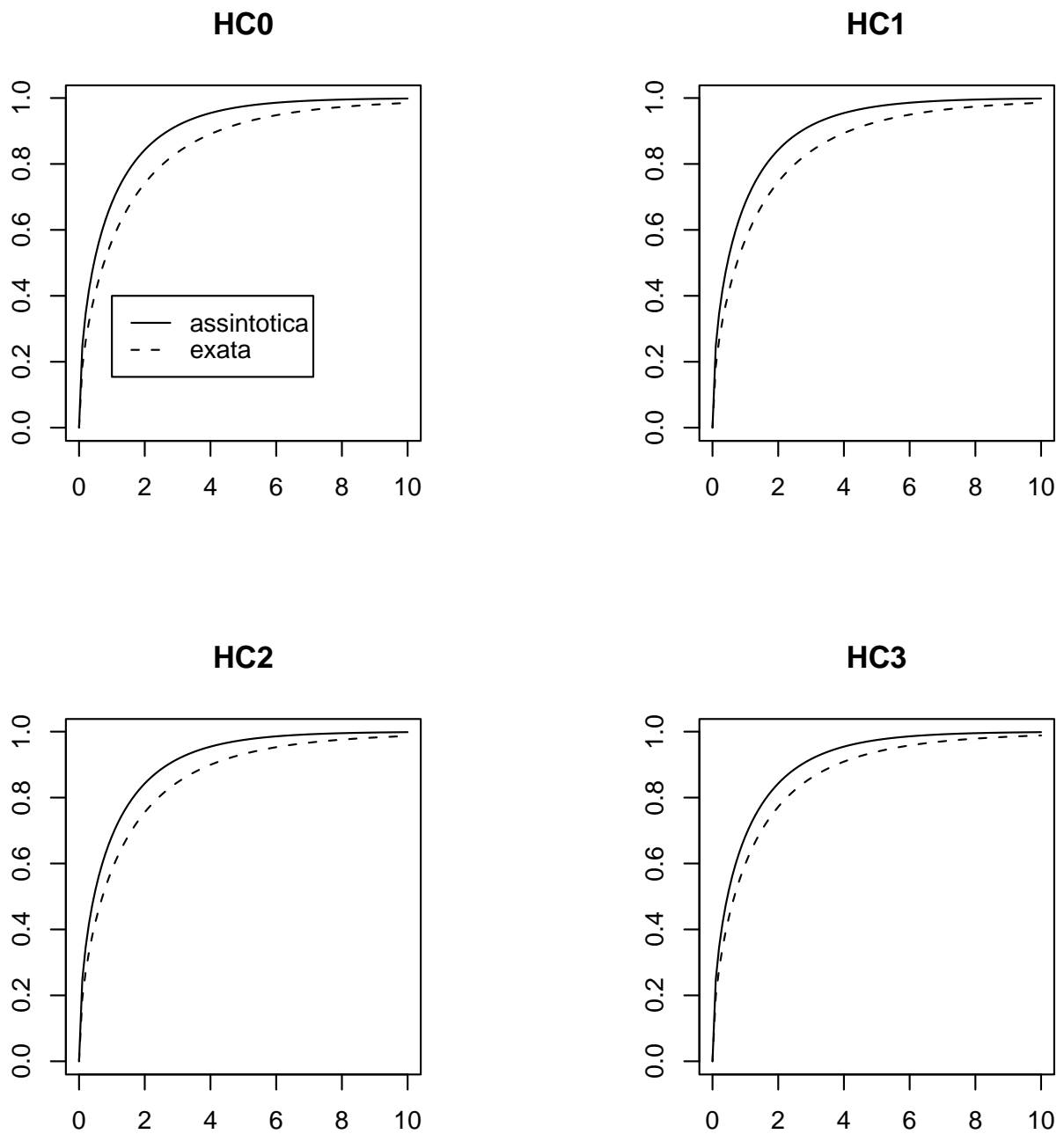


Figura 3.11. Distribuições exata e assintótica de t^2 , $n = 25$, $\alpha_1 = \alpha_2 = 0.415$, $\lambda = 101.513$, $x \sim t_3$, correção com duas iterações.

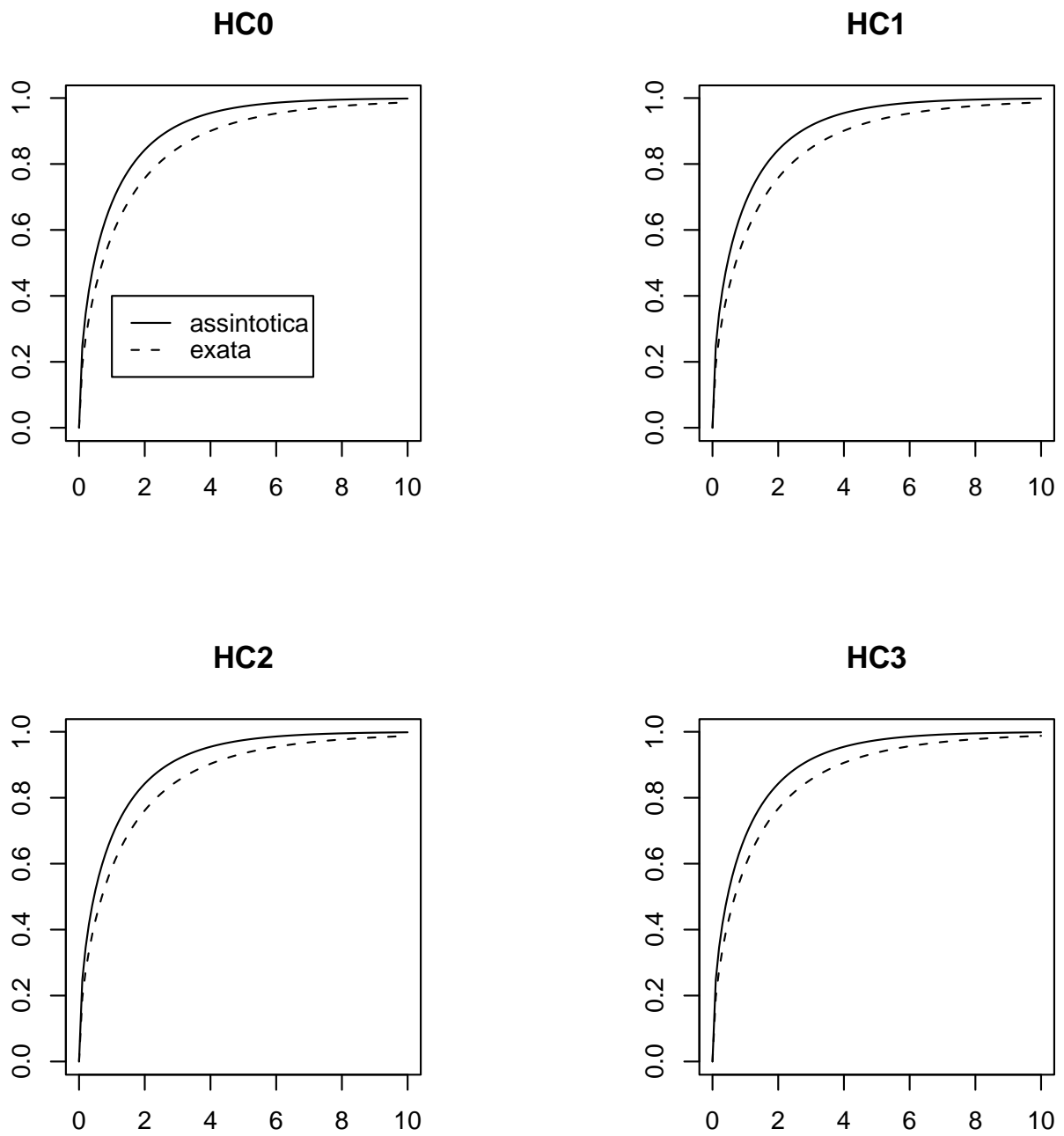


Figura 3.12. Distribuições exata e assintótica de t^2 , $n = 25$, $\alpha_1 = \alpha_2 = 0.415$, $\lambda = 101.513$, $x \sim t_3$, correção com cinco iterações.

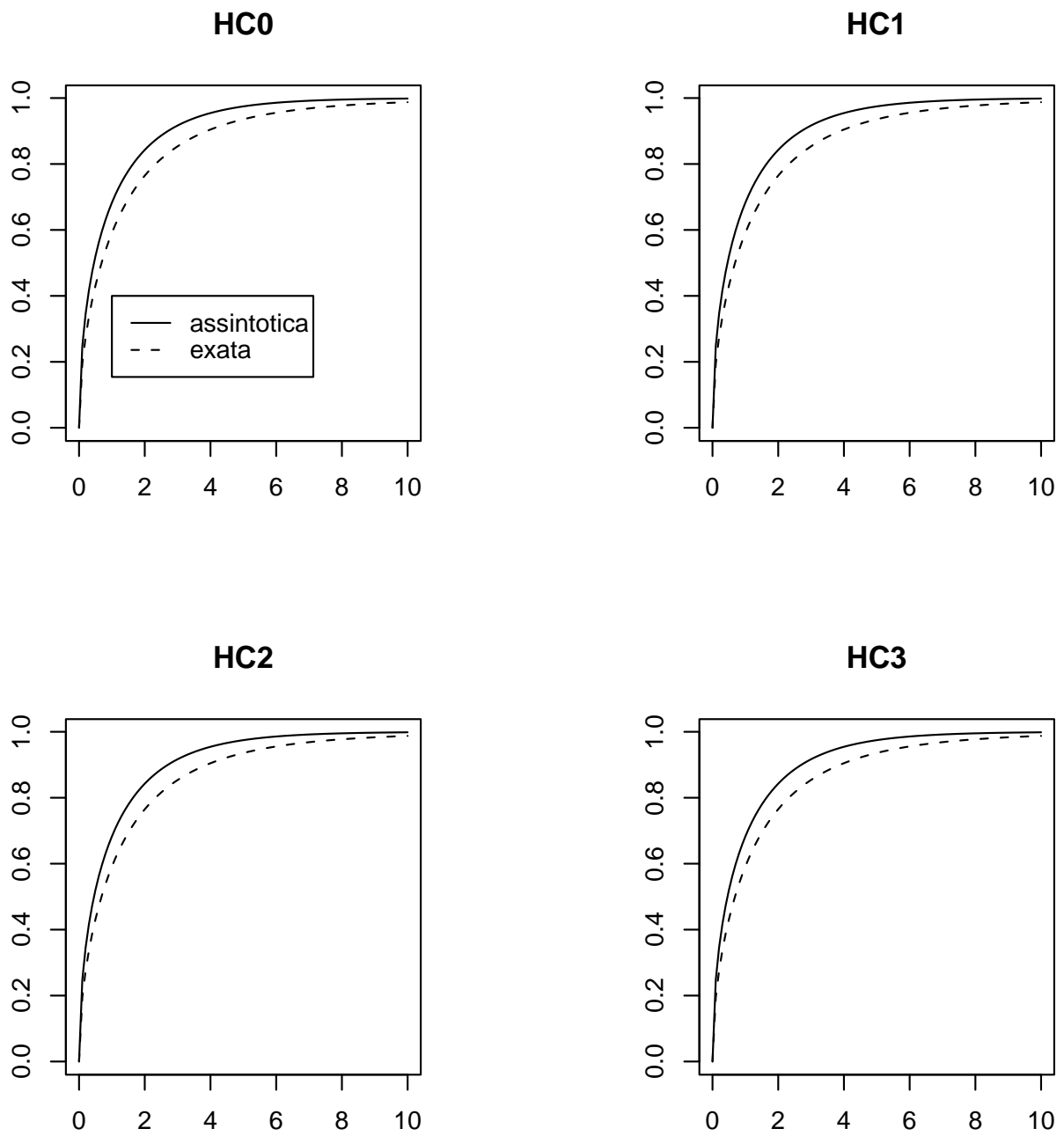


Figura 3.13. Distribuições exata e assintótica de t^2 , $n = 50$, $\alpha_1 = \alpha_2 = 0.415$, $\lambda = 101.513$, $x \sim t_3$, sem correção.

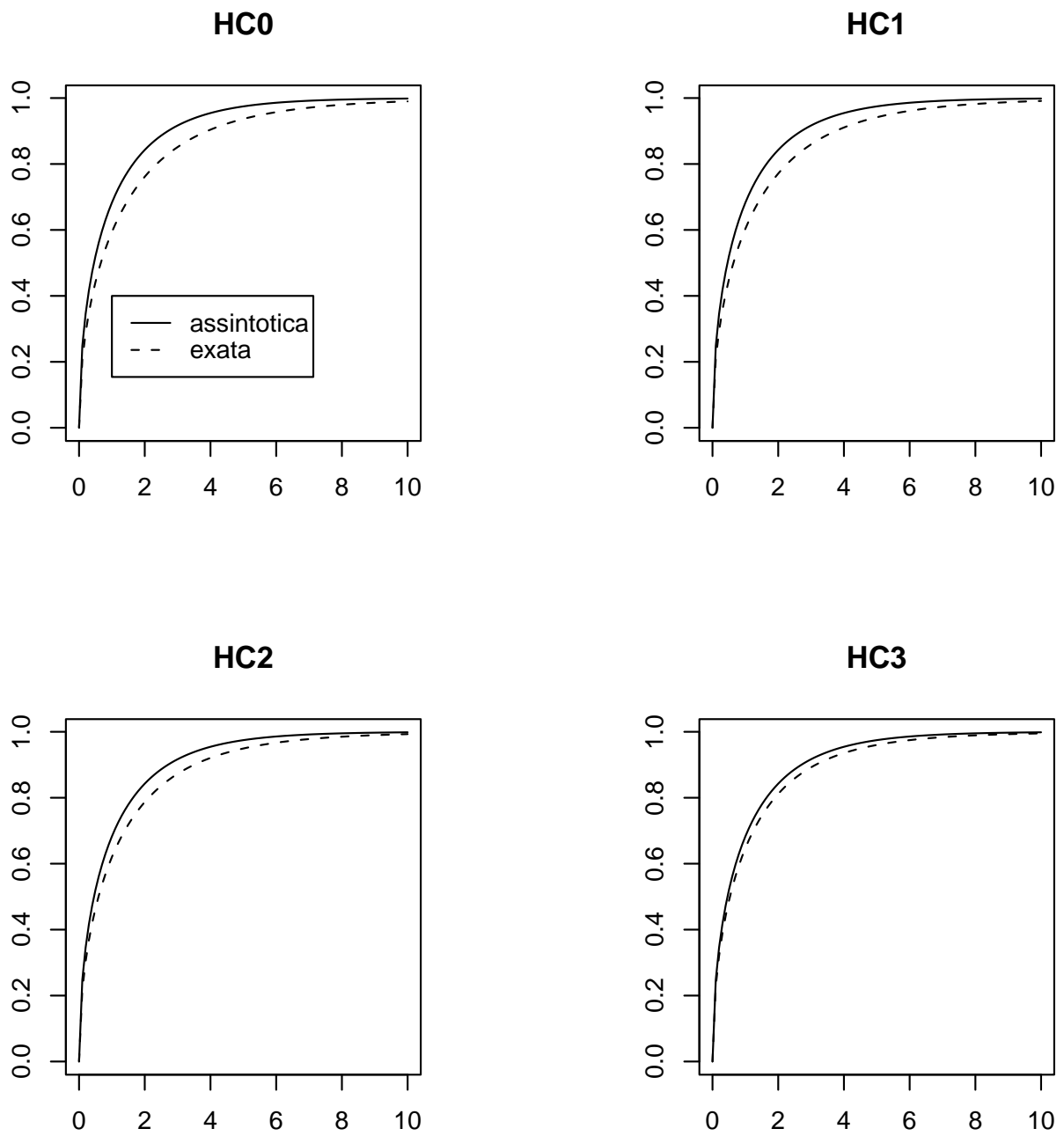


Figura 3.14. Distribuições exata e assintótica de t^2 , $n = 50$, $\alpha_1 = \alpha_2 = 0.415$, $\lambda = 101.513$, $x \sim t_3$, correção com uma iteração.

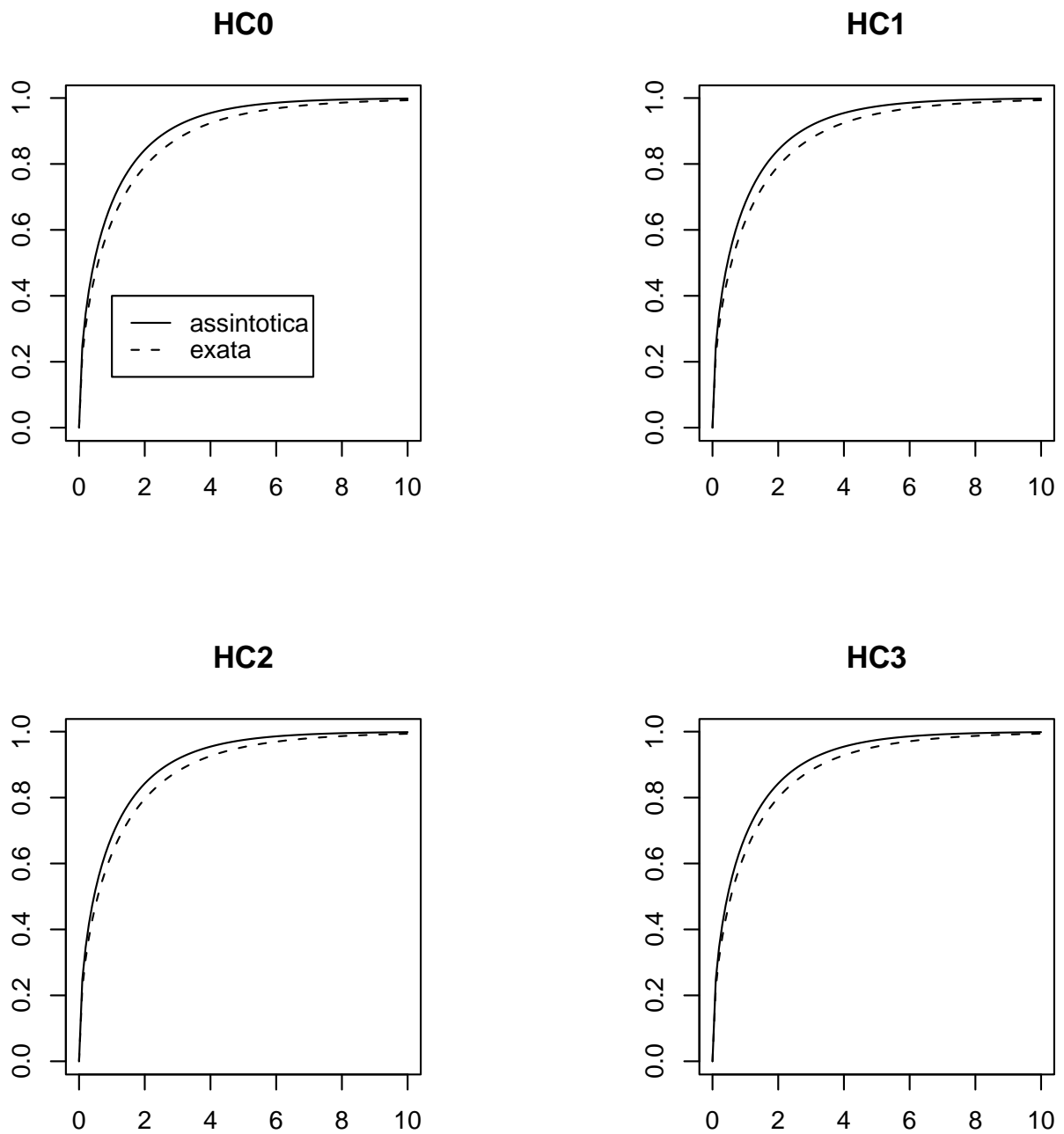


Figura 3.15. Distribuições exata e assintótica de t^2 , $n = 50$, $\alpha_1 = \alpha_2 = 0.415$, $\lambda = 101.513$, $x \sim t_3$, correção com duas iterações.

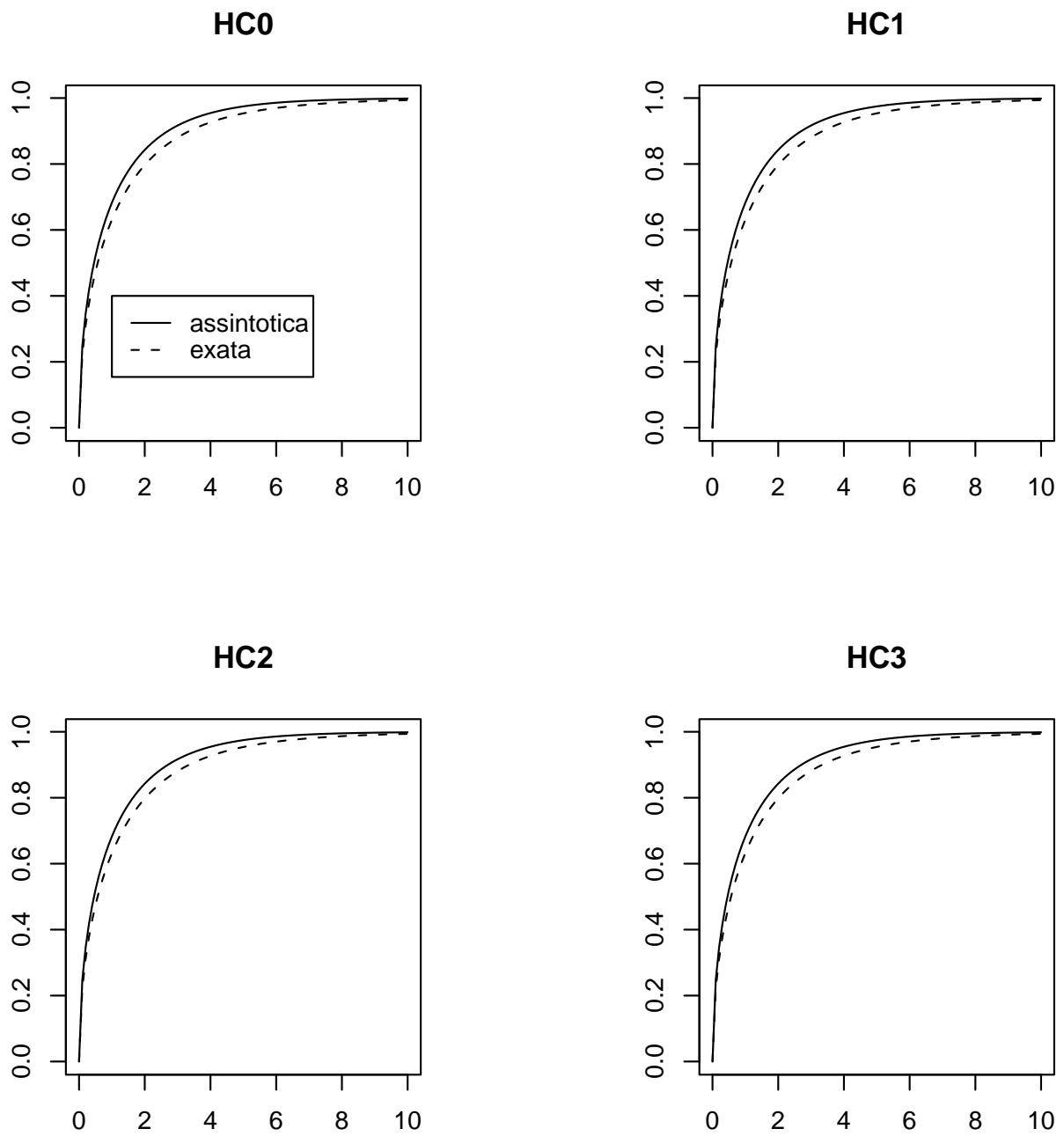


Figura 3.16. Distribuições exata e assintótica de t^2 , $n = 50$, $\alpha_1 = \alpha_2 = 0.415$, $\lambda = 101.513$, $x \sim t_3$, correção com cinco iterações.

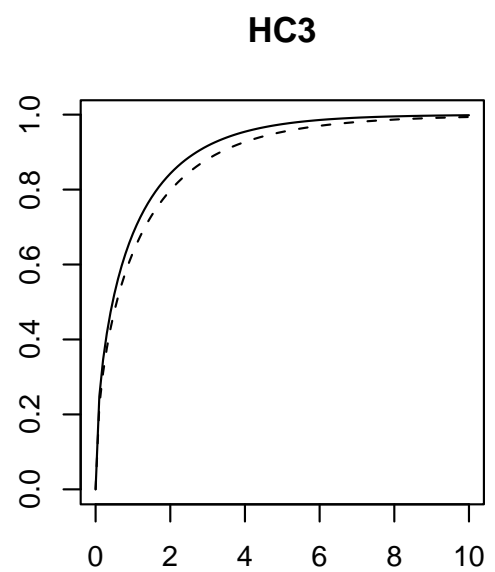
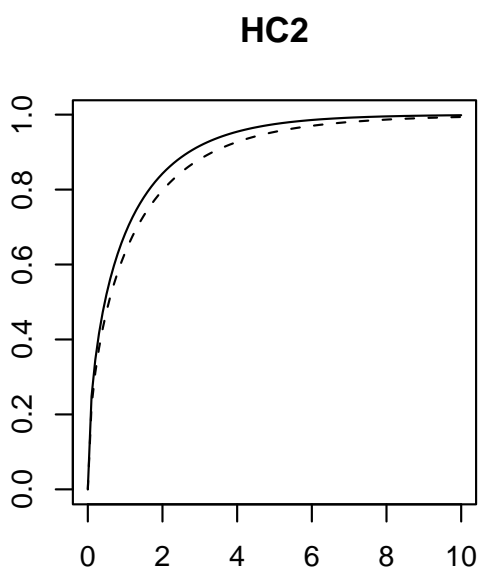
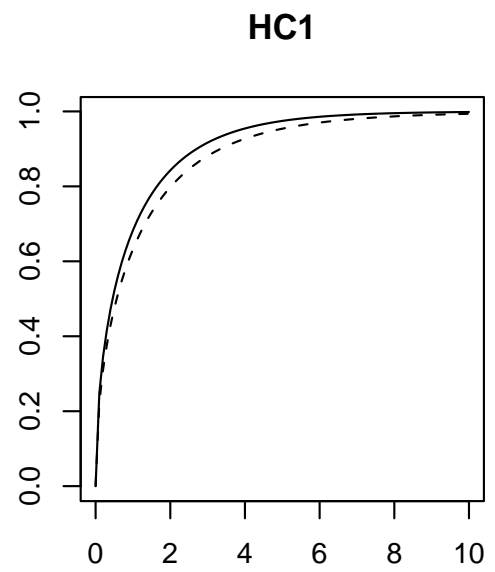
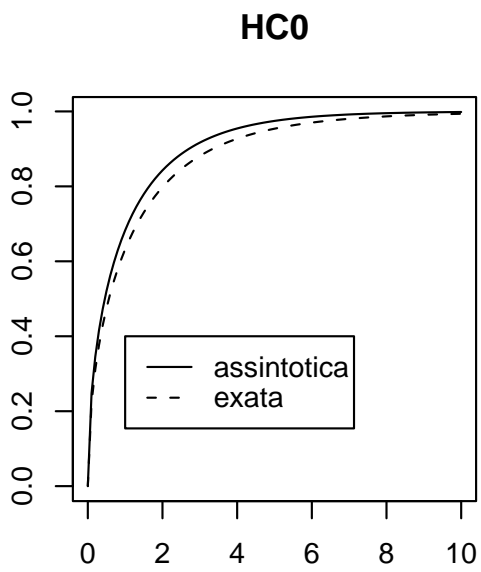


Figura 3.17. Distribuições exata e assintótica de t^2 , $n = 25$, $\alpha_1 = \alpha_2 = 0$, $\lambda = 1$, $x \sim \mathcal{U}(0, 1)$, sem correção.

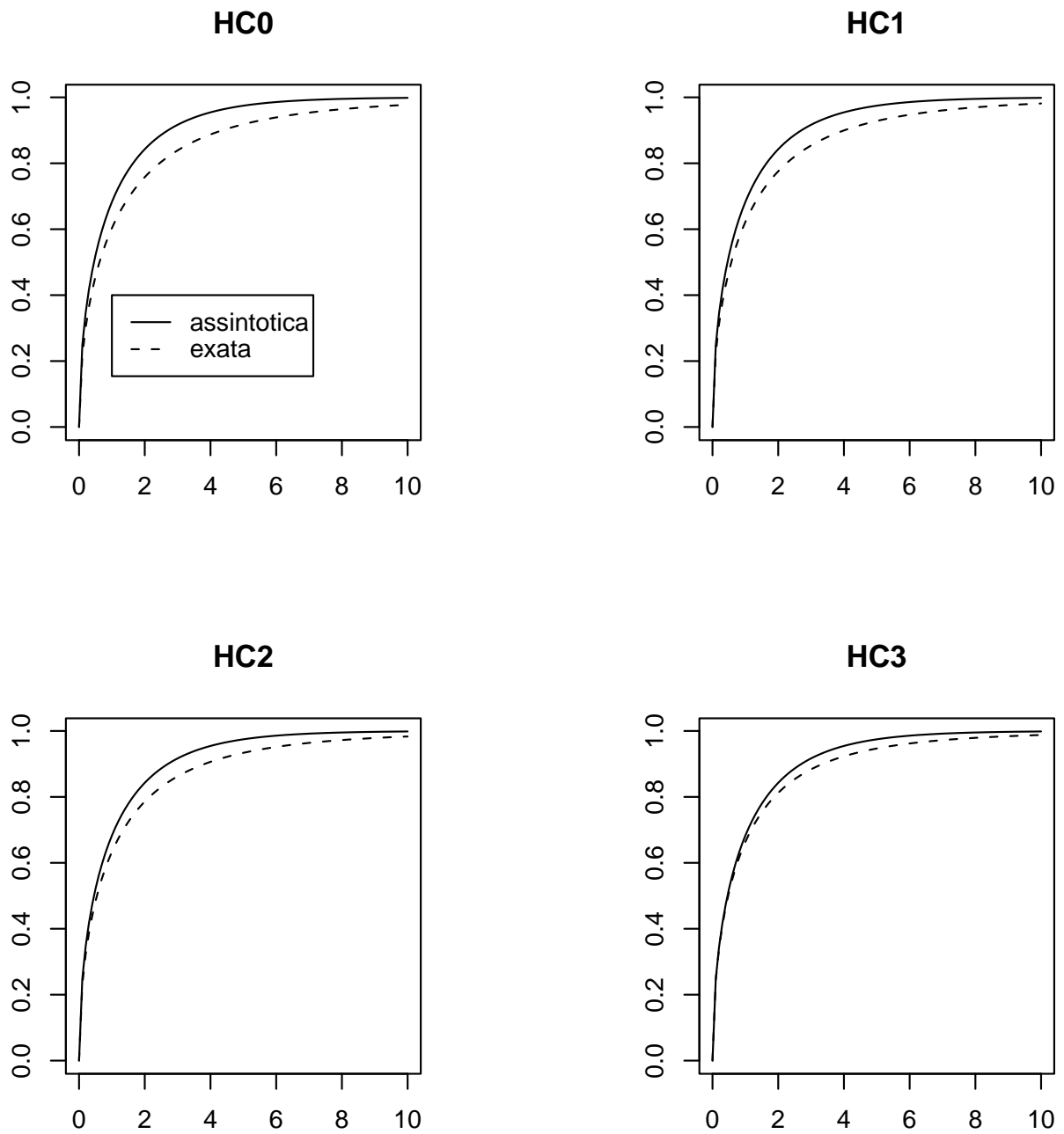


Figura 3.18. Distribuições exata e assintótica de t^2 , $n = 25$, $\alpha_1 = \alpha_2 = 0$, $\lambda = 1$, $x \sim \mathcal{U}(0, 1)$, correção com uma iteração.

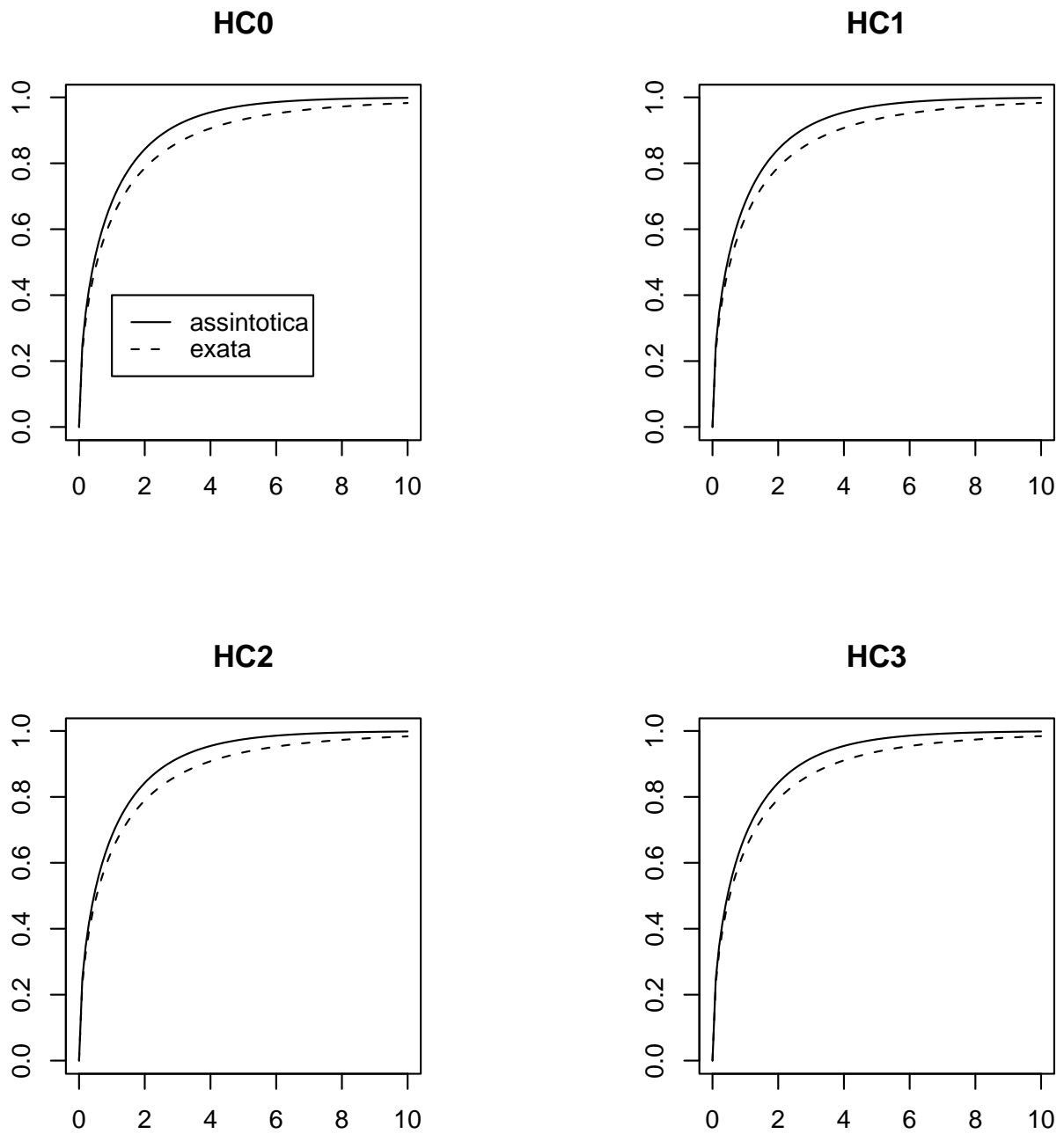


Figura 3.19. Distribuições exata e assintótica de t^2 , $n = 25$, $\alpha_1 = \alpha_2 = 0$, $\lambda = 1$, $x \sim \mathcal{U}(0, 1)$, correção com duas iterações.

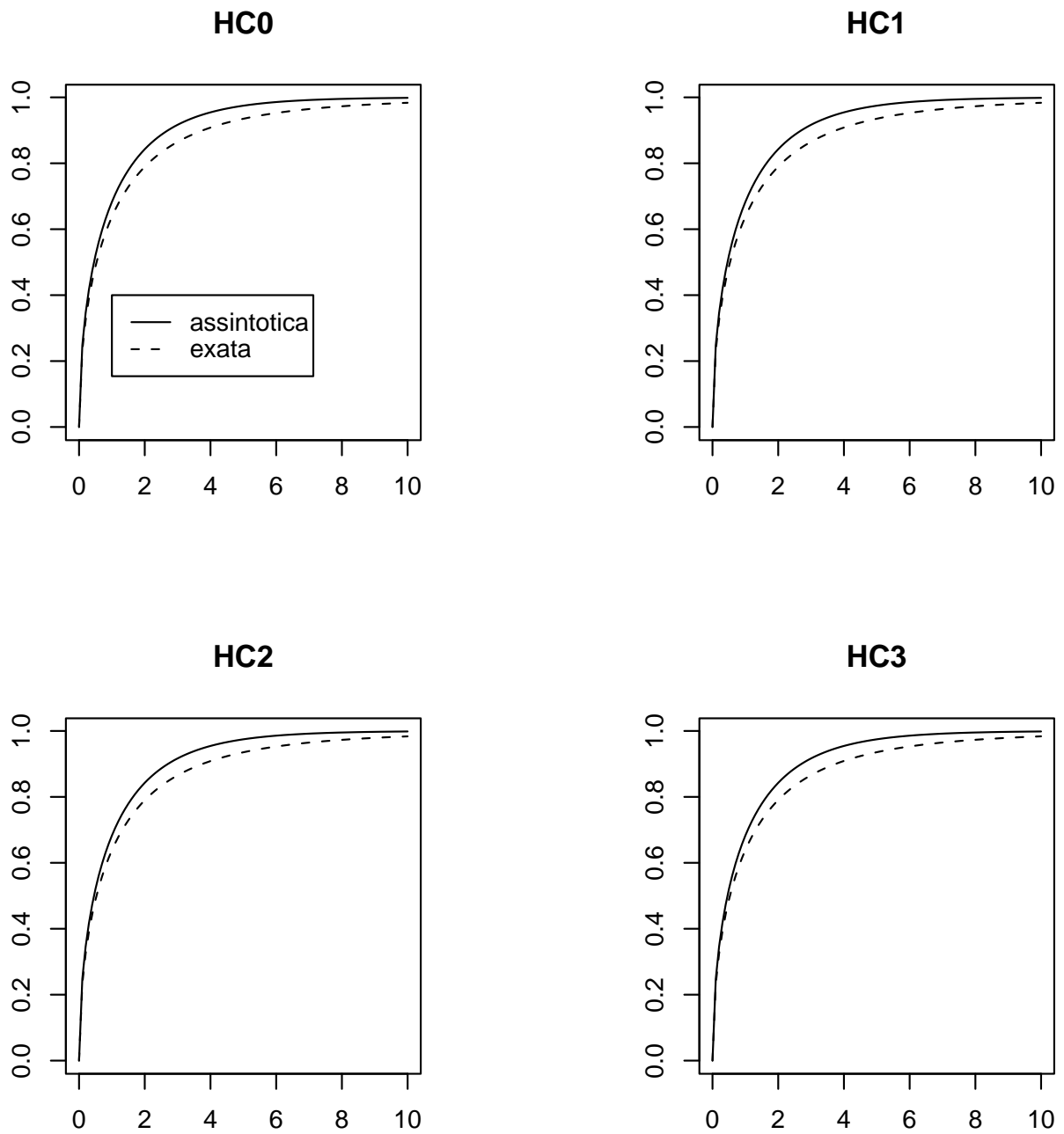


Figura 3.20. Distribuições exata e assintótica de t^2 , $n = 50$, $\alpha_1 = \alpha_2 = 0$, $\lambda = 1$, $x \sim \mathcal{U}(0, 1)$, sem correção.

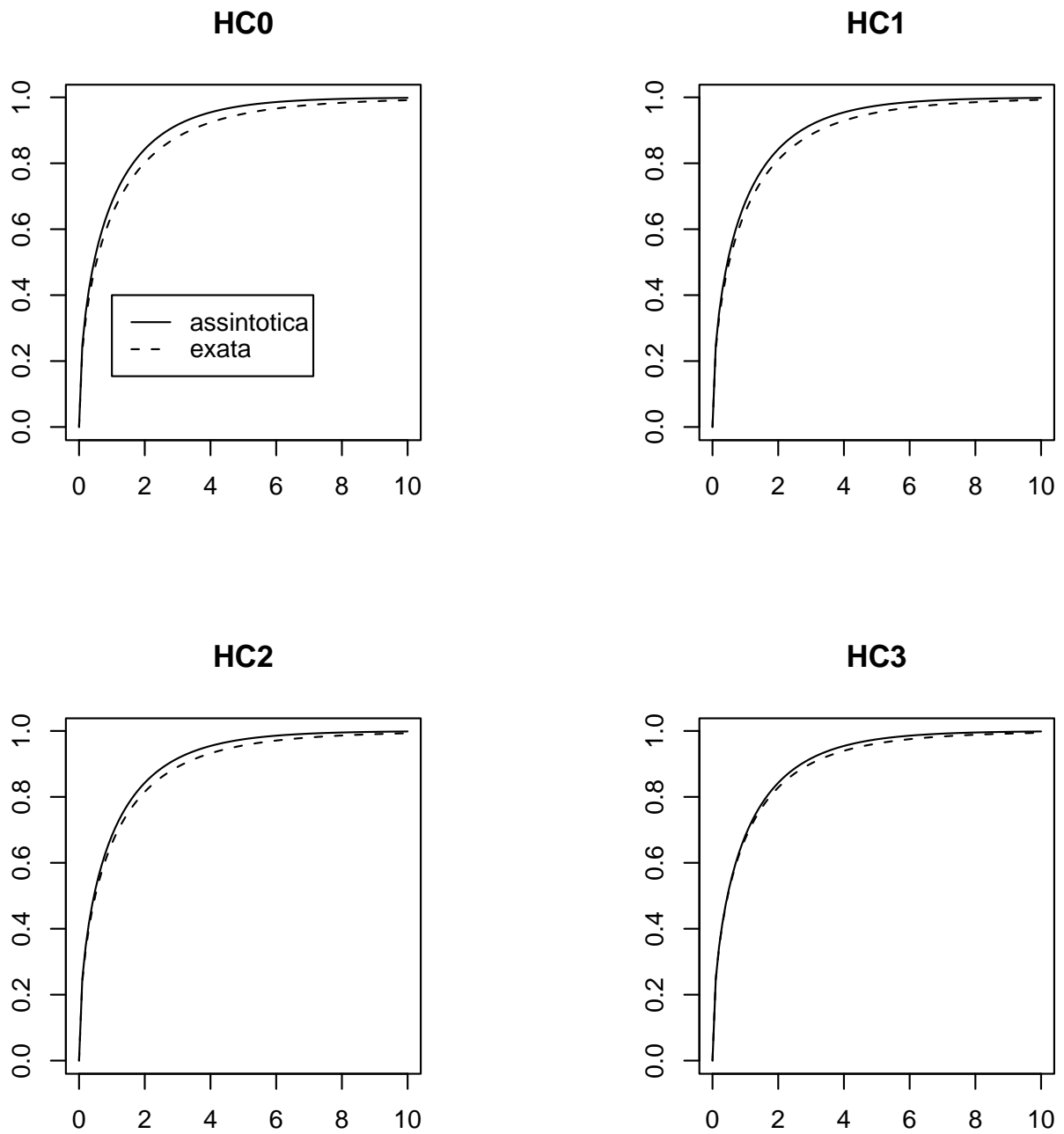


Figura 3.21. Distribuições exata e assintótica de t^2 , $n = 50$, $\alpha_1 = \alpha_2 = 0$, $\lambda = 1$, $x \sim \mathcal{U}(0, 1)$, correção com uma iteração.

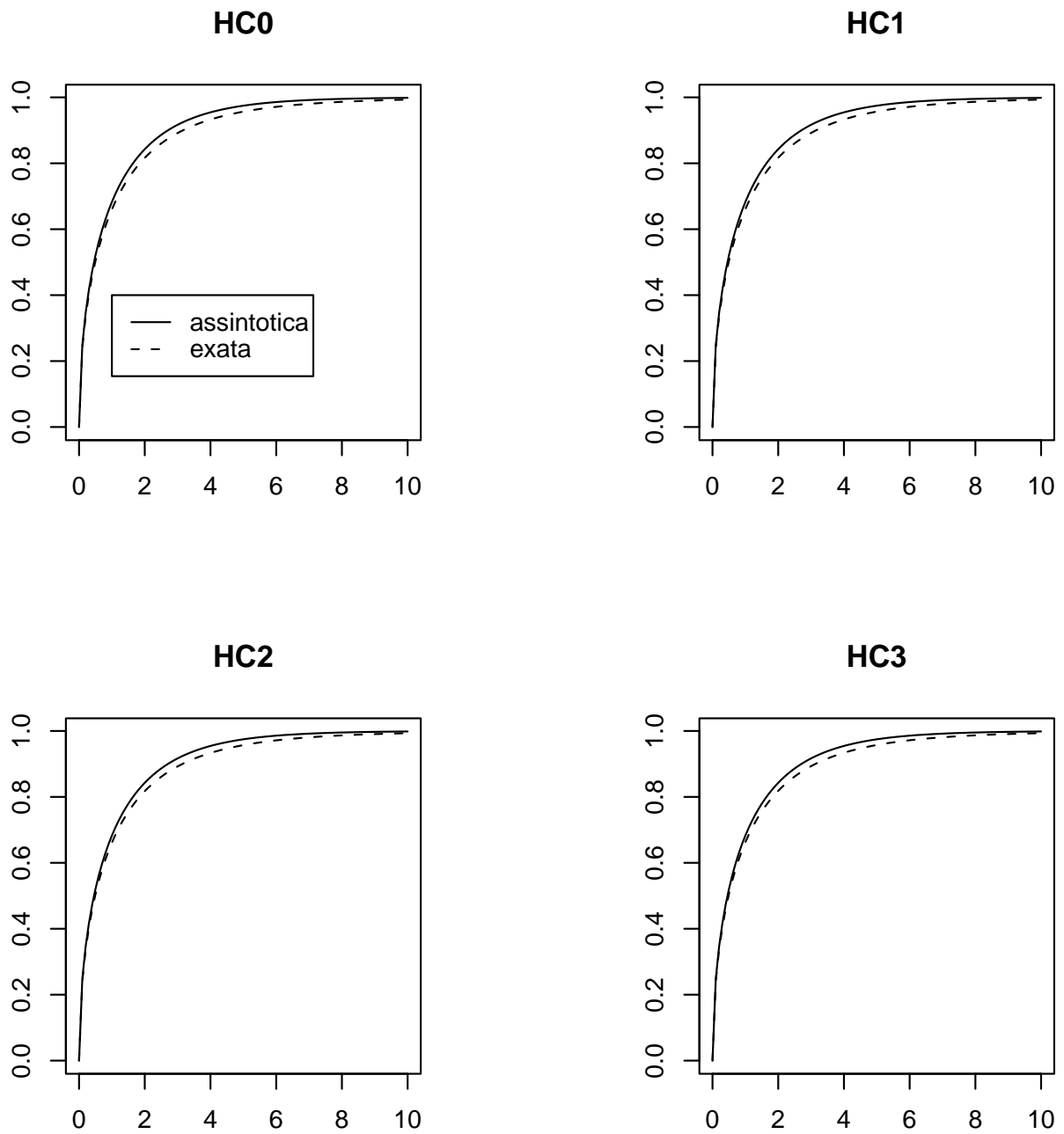


Figura 3.22. Distribuições exata e assintótica de t^2 , $n = 50$, $\alpha_1 = \alpha_2 = 0$, $\lambda = 1$, $x \sim \mathcal{U}(0, 1)$, correção com duas iterações.

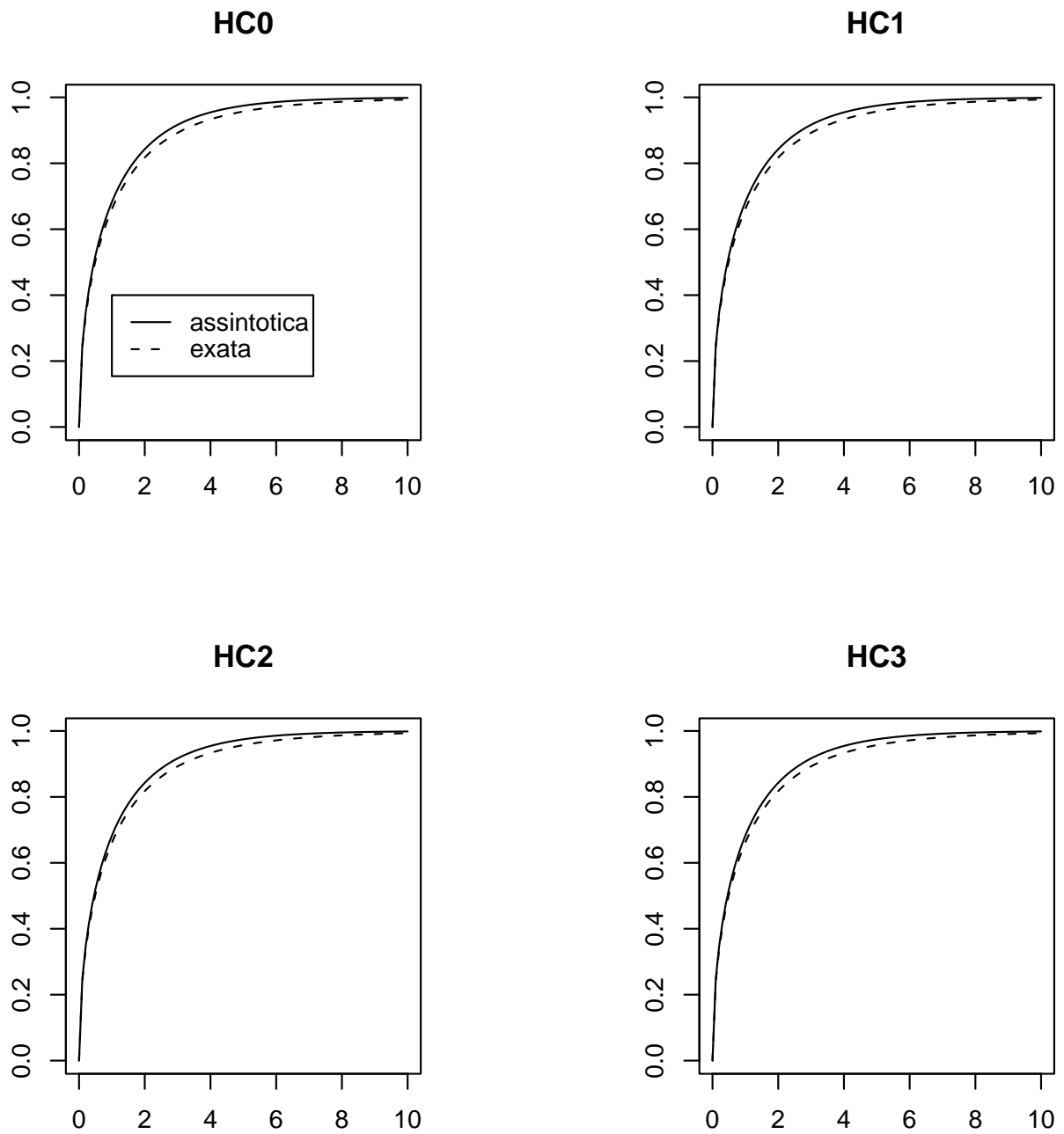


Figura 3.23. Distribuições exata e assintótica de t^2 , $n = 25$, $\alpha_1 = \alpha_2 = 0$, $\lambda = 1$, $x \sim t_3$, sem correção.

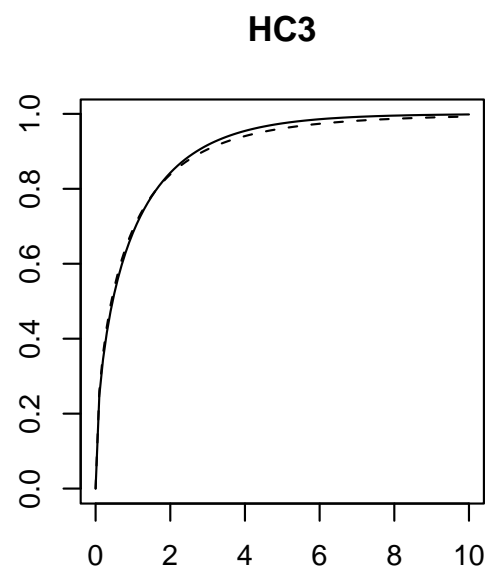
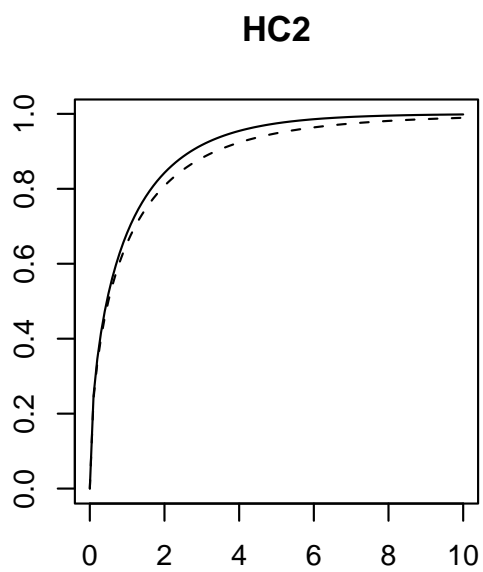
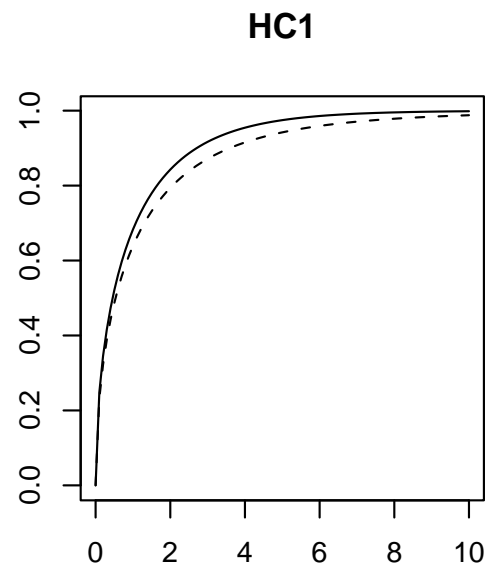
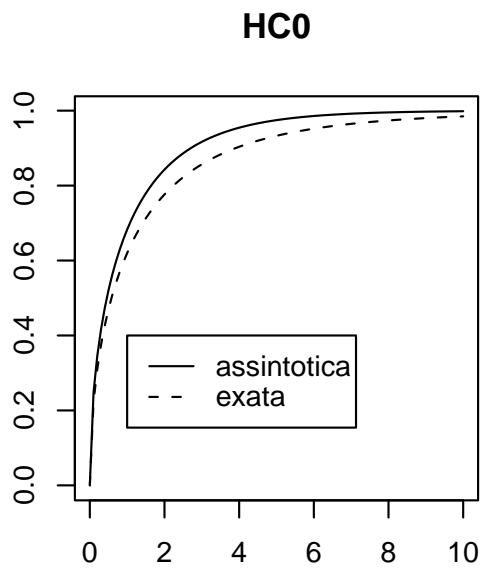


Figura 3.24. Distribuições exata e assintótica de t^2 , $n = 25$, $\alpha_1 = \alpha_2 = 0$, $\lambda = 1$, $x \sim t_3$, correção com uma iteração.

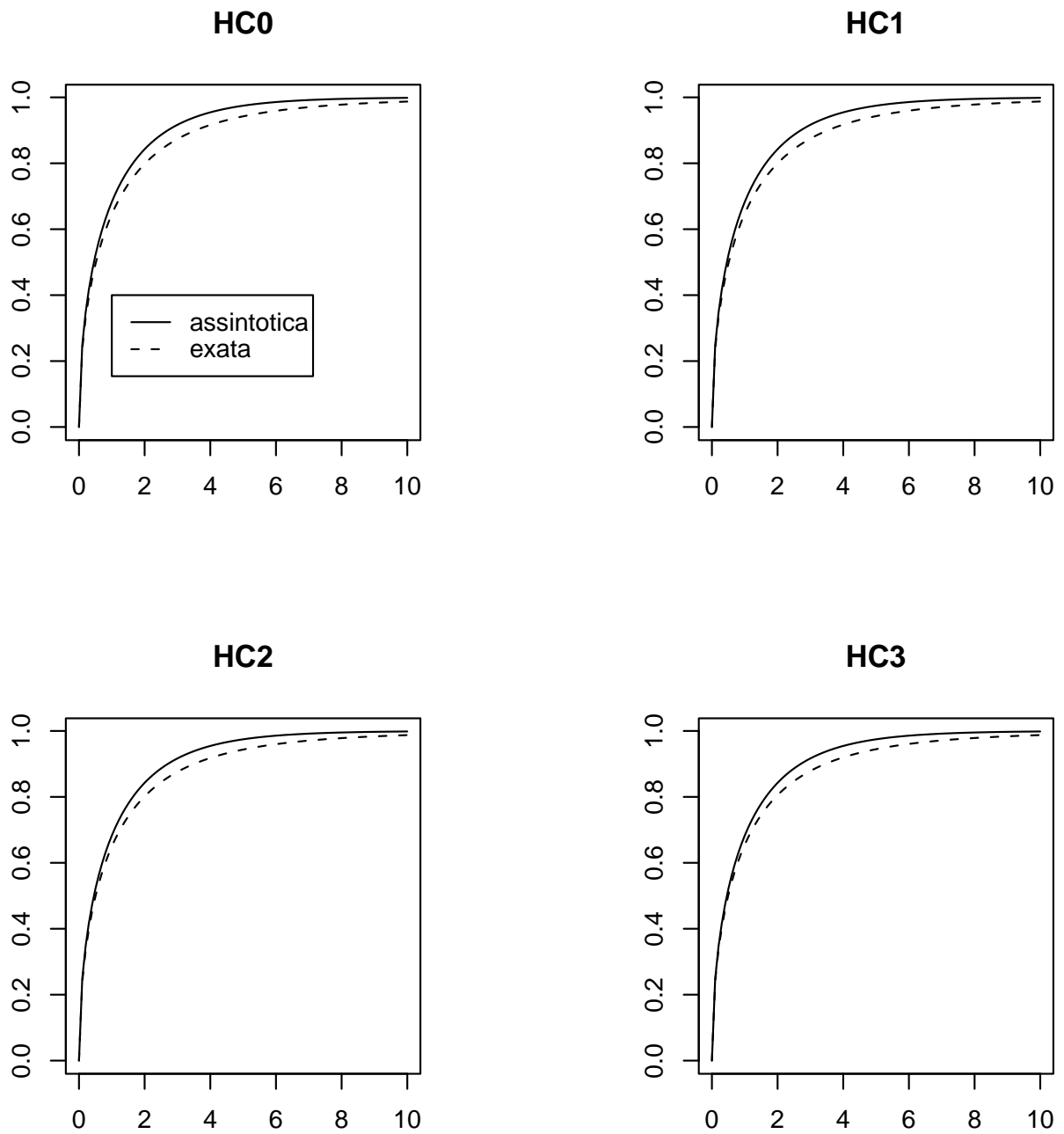


Figura 3.25. Distribuições exata e assintótica de t^2 , $n = 25$, $\alpha_1 = \alpha_2 = 0$, $\lambda = 1$, $x \sim t_3$, correção com duas iterações.

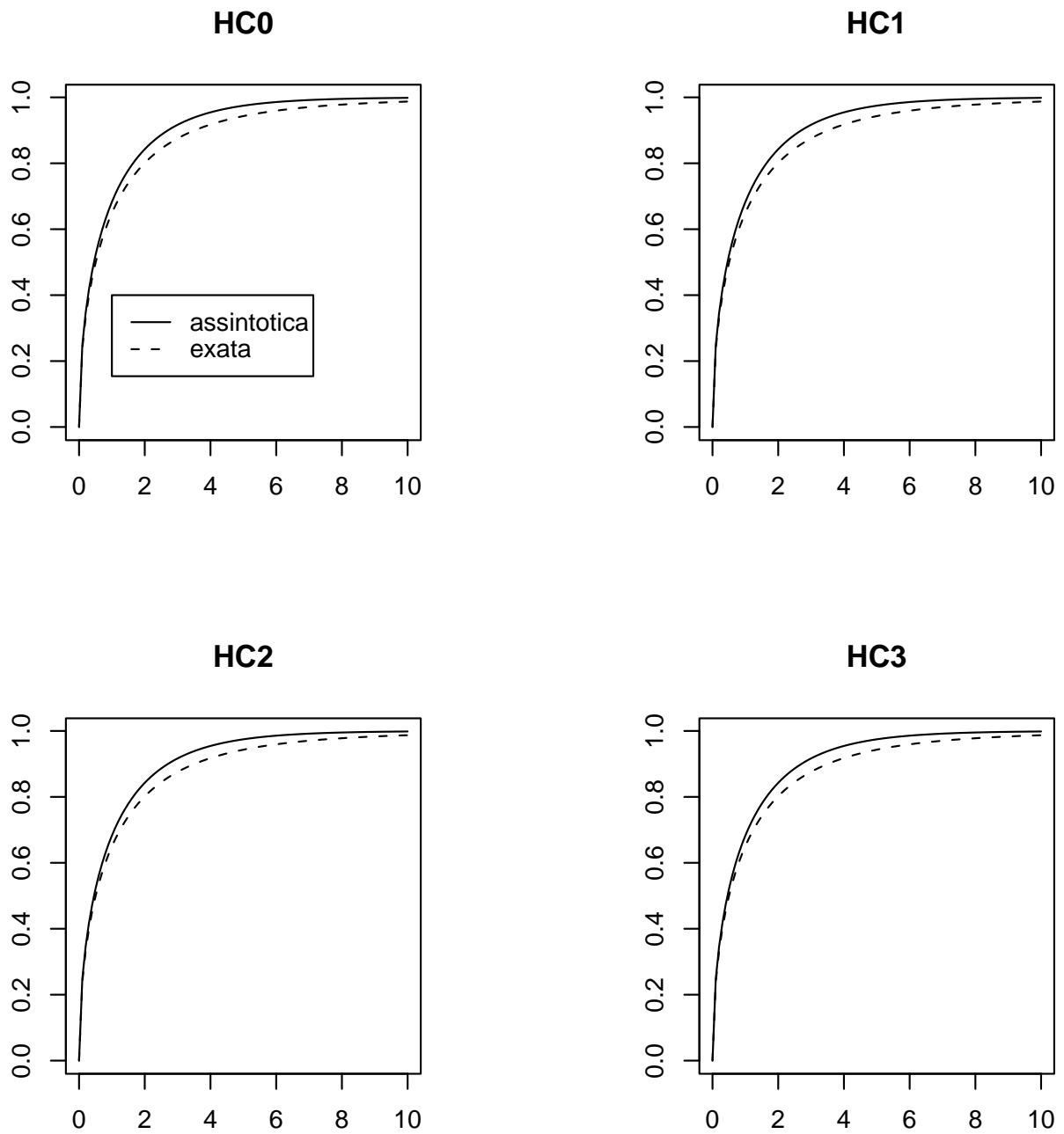


Figura 3.26. Distribuições exata e assintótica de t^2 , $n = 25$, $\alpha_1 = \alpha_2 = 0$, $\lambda = 1$, $x \sim t_3$, correção com dez iterações.

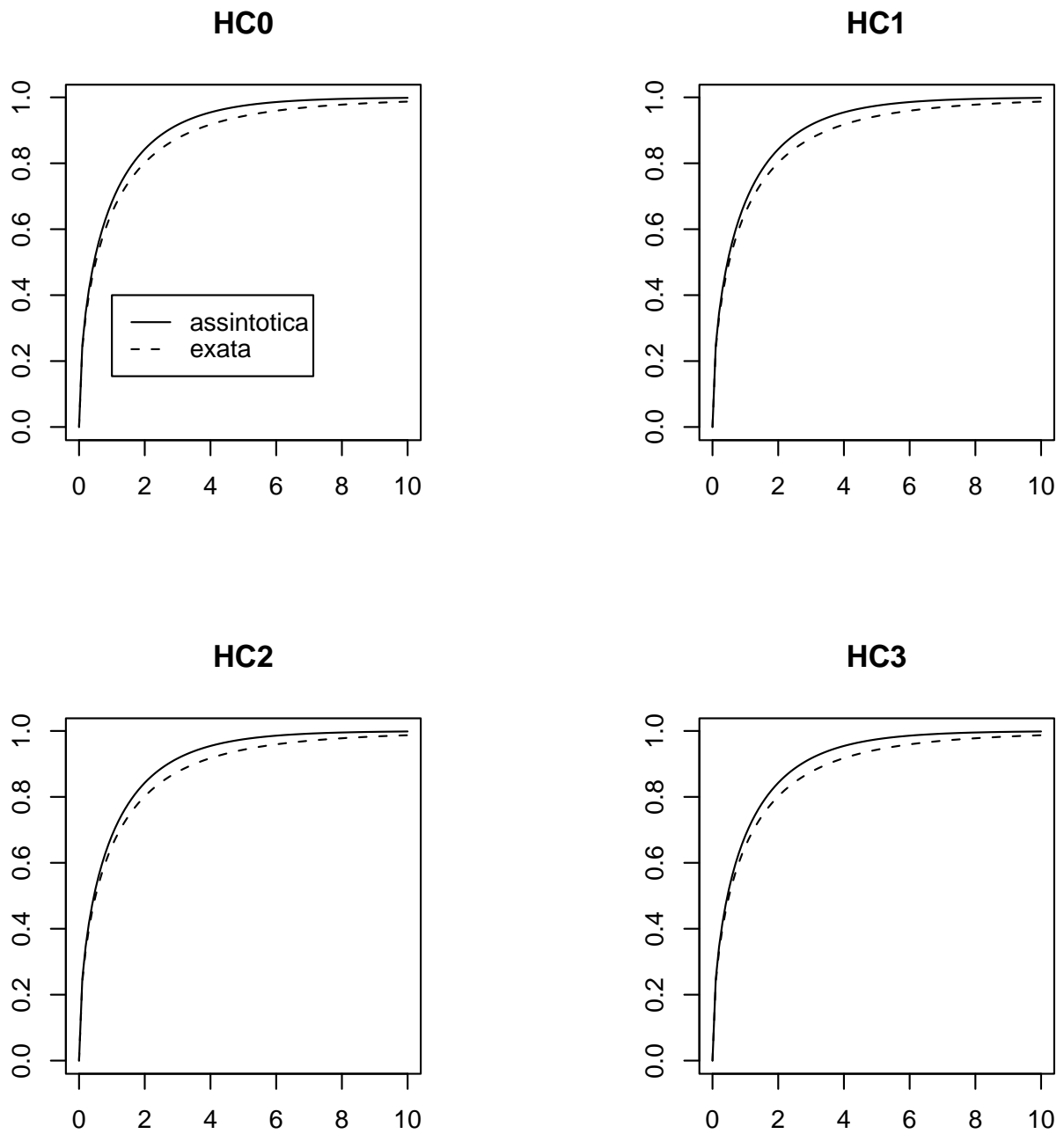


Figura 3.27. Distribuições exata e assintótica de t^2 , $n = 50$, $\alpha_1 = \alpha_2 = 0$, $\lambda = 1$, $x \sim t_3$, sem correção.

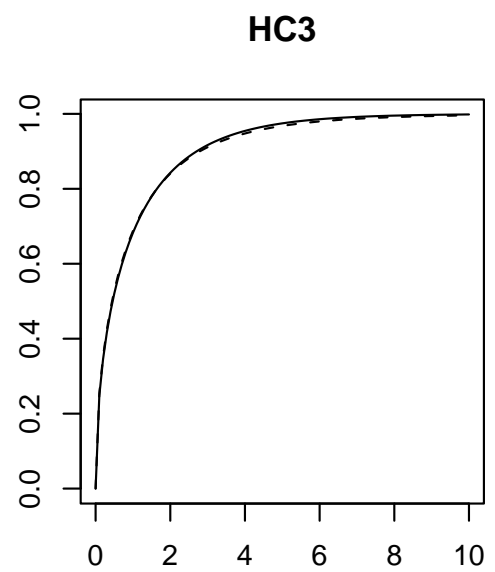
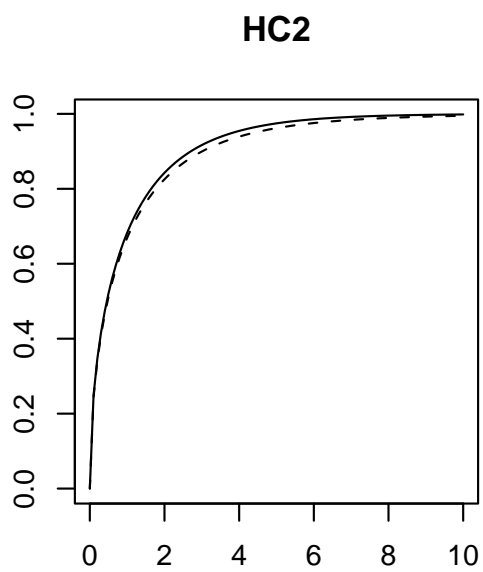
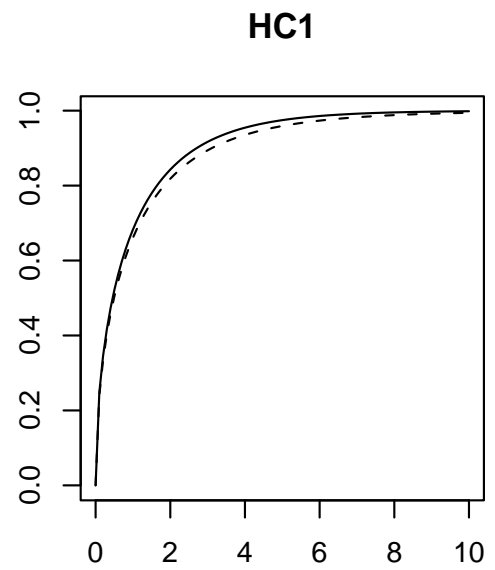
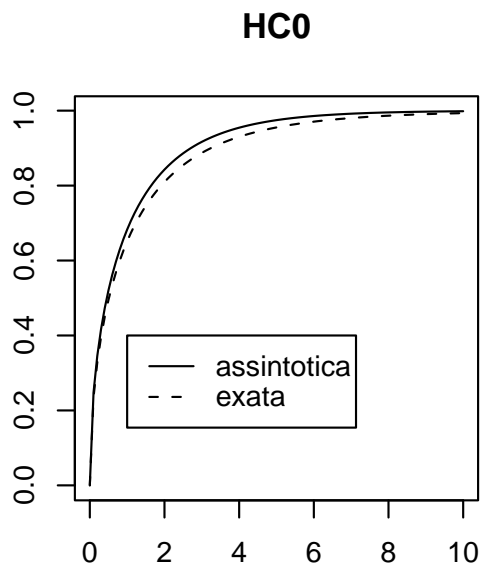
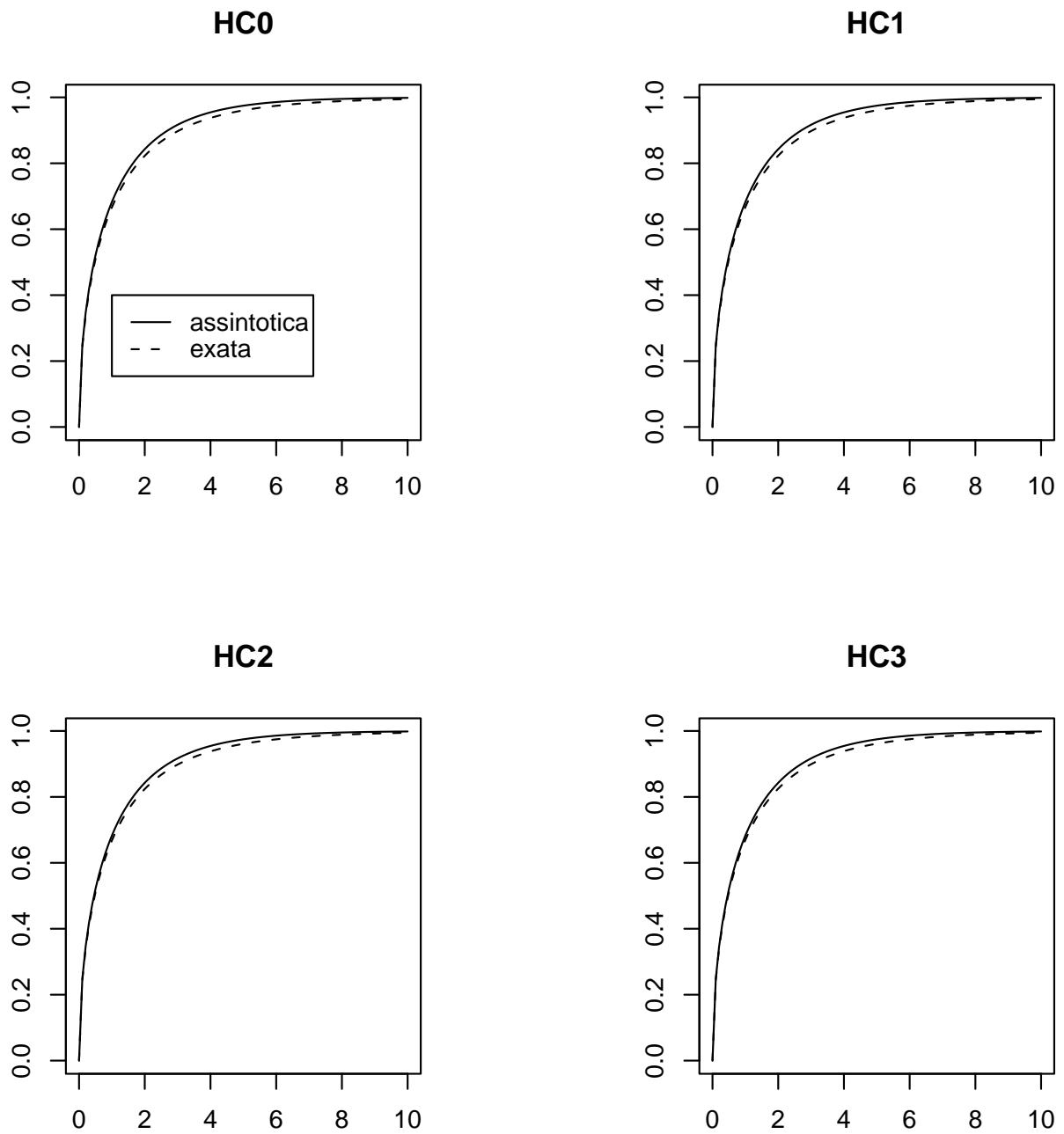


Figura 3.28. Distribuições exata e assintótica de t^2 , $n = 50$, $\alpha_1 = \alpha_2 = 0$, $\lambda = 1$, $x \sim t_3$, correção com uma iteração.



Referências

- [1] Box, G.E.P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification. *Annals of Mathematical Statistics*, 25, 290–302.
- [2] Chesher, A. & Austin, G. (1991). The finite-sample distributions of heteroskedasticity robust Wald statistics. *Journal of Econometrics*, 47, 153–173.
- [3] Chesher, A. & Jewitt, I. (1987). The bias of a heteroskedasticity consistent covariance matrix estimator. *Econometrica*, 55, 1217–1222.
- [4] Cribari-Neto, F. (1997). Econometric programming environments: GAUSS, Ox and S-PLUS. *Journal of Applied Econometrics*, 12, 77–89.
- [5] Cribari-Neto, F., Ferrari, S.L.P. & Cordeiro, G.M. (2000). Improved heteroscedasticity-consistent covariance matrix estimators. *Biometrika*, 87, 907–918.
- [6] Cribari-Neto, F. & Galvão, N.M.S. (2002). A class of improved heteroskedasticity-consistent covariance matrix estimators. Artigo submetido para publicação.
- [7] Cribari-Neto, F. & Zarkos, S.G. (1999). Bootstrap methods for heteroskedastic regression models: evidence on estimation and testing. *Econometric Reviews*, 18, 211–228.
- [8] Davidson, R. & MacKinnon, J.G. (1993). *Estimation and Inference in Econometrics*. New York: Oxford University Press.
- [9] Doornik, J.A. (2001). *Ox: an Object-oriented Matrix Programming Language*, 4^a ed. Londres: Timberlake Consultants e Oxford: <http://www.nuff.ox.ac.uk/Users/Doornik>.
- [10] Farebrother, R.W. (1990). The distribution of a quadratic form in normal variables. *Applied Statistics*, 39, 294–309.
- [11] Galvão, N.M.S. (2000). *Estimação Consistente de Matrizes de Covariância sob Heterocedasticidade de Forma Desconhecida*. Dissertação de Mestrado, Programa de Pós-graduação em Estatística, Universidade Federal de Pernambuco.
- [12] Gil-Pelaez, J. (1951). Note on the inversion theorem. *Biometrika*, 38, 481–482.
- [13] Greene, W.H. (1997). *Econometric Analysis*, 3^a ed. Upper Saddle River: Prentice Hall.
- [14] Harvey, A.C. (1976). Estimating regression models with multiplicative heteroscedasticity. *Econometrica*, 44, 461–465.
- [15] Hinkley, D.V. (1977). Jackknifing in unbalanced situations. *Technometrics*, 19, 285–292.
- [16] Horn, S.D., Horn, R.A. & Duncan, D.B. (1975). Estimating heteroskedastic variances in linear models. *Journal of the American Statistical Association*, 70, 380–385.

- [17] Imhof, J.P. (1961). Computing the distribution of a quadratic form in normal variates. *Biometrika*, 48, 419–426.
- [18] Judge, G.C, Hill, R.C., Griffiths, W.E., Lutkepohl, H. & Lee, T.C. (1988). *Introduction to the Theory and Practice of Econometrics*, 2nd ed. New York: Wiley.
- [19] Knuth, D.E. (1986). *The T_EXbook*. New York: Addison–Wesley.
- [20] Long, J.S. & Ervin, L.H. (2000). Using heteroskedasticity-consistent standard errors in the linear regression model. *The American Statistician*, 54, 217–224.
- [21] MacKinnon, J.G. & White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite-sample properties. *Journal of Econometrics*, 29, 305–325.
- [22] Scheffé, H. (1959). *The Analysis of Variance*. New York: Wiley.
- [23] Searle, S.R. (1977). *Linear Models*. New York: Wiley.
- [24] Seber, G.A.F. (1977). *Linear Regression Analysis*. New York: Wiley.
- [25] White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48, 817–838.
- [26] White, H. (1984). *Asymptotic Theory for Econometricians*. Orlando: Academic Press.